TRB

TRANSPORTATION RESEARCH RECORD

Journal of the Transportation Research Board, No. 2175

Travel Forecasting 2010

VOLUME 1

TRANSPORTATION RESEARCH BOARD
OF THE NATIONAL ACADEMIES

TRANSPORTATION RESEARCH BOARD
OF THE NATIONAL ACADEMIES

TRANSPORTATION RESEARCH RECORDS, which are published throughout the year, consist of collections of papers on specific transportation modes and subject areas. Each Record is classified according to the subscriber category or categories covered by the papers published in that volume. The views expressed in papers published in the Transportation Research Record series are those of the authors and do not necessarily reflect the views of the peer review committee(s), the Transportation Research Board, the National Research Council, or the sponsors of TRB activities. The Transportation Research Board does not endorse products or manufacturers; trade and manufacturers' names may appear in a Record paper only if they are considered essential.

PEER REVIEW OF PAPERS: All papers published in the Transportation Research Record series have been reviewed and accepted for publication through the Transportation Research Board's peer review process established according to procedures approved by the Governing Board of the National Research Council. Papers are refereed by the TRB standing committees identified on page ii of each Record. Reviewers are selected from among committee members and other outside experts. The Transportation Research Board requires a minimum of three reviews; a decision is based on reviewer comments and resultant author revision. For details about the peer review process, see the information on the inside back cover.

THE TRANSPORTATION RESEARCH RECORD PUBLICATION BOARD, comprising a cross section of transportation disciplines and with equal representation from the academic and practitioner communities, assures that the quality of the *Transportation Research Record: Journal of the Transportation Research Board* and the TRB paper peer review process is consistent with that of a peer-reviewed scientific journal. Members from the academic community make decisions on granting tenure; all academic and practitioner members have authored papers published in peer-reviewed journals and all have participated in the TRB peer review process.

TRANSPORTATION RESEARCH BOARD PUBLICATIONS may be ordered directly from the TRB Business Office, through the Internet at www.TRB.org, or by annual subscription through organizational or individual affiliation with TRB. Affiliates and library subscribers are eligible for substantial discounts. For further information, contact the Transportation Research Board Business Office, 500 Fifth Street, NW, Washington, DC 20001 (telephone 202-334-3213; fax 202-334-2519; or e-mail TRBsales@nas.edu).

TRANSPORTATION RESEARCH RECORD PAPERS ONLINE: The TRR Journal Online website provides electronic access to the full text of papers that have been published in the Transportation Research Record series since 1996. The site is updated as new Record papers become available. To search abstracts and to find subscription and pricing information, go to www.TRB.org/TRROnline.

TRB

TRANSPORTATION RESEARCH RECORD

Journal of the Transportation Research Board, No. 2175

Travel Forecasting 2010

VOLUME 1

A Peer-Reviewed Publication

TRANSPORTATION RESEARCH BOARD
OF THE NATIONAL ACADEMIES

Washington, D.C.
2010

www.TRB.org

Transportation Research Record 2175
ISSN 0361-1981
ISBN 978-0-309-16051-3

Subscriber Categories
Highways; public transportation; pedestrians and bicyclists; planning and forecasting; passenger transportation; economics; environment; data and information technology

Printed in the United States of America

Transportation Research Board Staff
Kimberly M. Fisher, Associate Division Director, Transportation Planning
Freda R. Morgan, Senior Program Associate
Mary Kissi, Senior Program Associate

Publications Office
Diane LeBlanc Solometo, Editor; Elizabeth C. Horowitz, Production Editor; Jennifer Correro, Proofreader; Claudia Sauls, Manuscript Preparer

Ann E. Petty, Managing Editor; Juanita Green, Production Manager; Phyllis Barber, Publishing Administrator; Jennifer J. Weeks, Manuscript Preparation Manager

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. On the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Charles M. Vest is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, on its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both the Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Charles M. Vest are chair and vice chair, respectively, of the National Research Council.

The **Transportation Research Board** is one of six major divisions of the National Research Council. The mission of the Transportation Research Board is to provide leadership in transportation innovation and progress through research and information exchange, conducted within a setting that is objective, interdisciplinary, and multimodal. The Board's varied activities annually engage about 7,000 engineers, scientists, and other transportation researchers and practitioners from the public and private sectors and academia, all of whom contribute their expertise in the public interest. The program is supported by state transportation departments, federal agencies including the component administrations of the U.S. Department of Transportation, and other organizations and individuals interested in the development of transportation. **www.TRB.org**

www.national-academies.org

TRB SPONSORS*

Transportation Departments of the 50 States and the District of Columbia

Federal Government
U.S. Department of Transportation
- Federal Aviation Administration
- Federal Highway Administration
- Federal Motor Carrier Safety Administration
- Federal Railroad Administration
- Federal Transit Administration
- National Highway Traffic Safety Administration
- Research and Innovative Technology Administration

Bureau of Indian Affairs
Science and Technology Directorate,
U.S. Department of Homeland Security
U.S. Army Corps of Engineers
U.S. Coast Guard

Nongovernmental Organizations
American Association of State Highway and Transportation Officials
American Public Transportation Association
American Transportation Research Institute
Association of American Railroads

---

*As of November 2010.

TRANSPORTATION RESEARCH RECORD

Journal of the Transportation Research Board, No. 2175

Contents

Foreword

The 2010 series of the *Transportation Research Record: Journal of the Transportation Research Board* consists of approximately 900 papers selected from 3,700 submissions after rigorous peer review. The peer review for each paper published in this volume was coordinated by the committee acknowledged at the end of the text; members of the reviewing committees for the papers in this volume are listed on page ii.

Additional information about the *Transportation Research Record: Journal of the Transportation Research Board* series and the peer review process appears on the inside back cover. TRB appreciates the interest shown by authors in offering their papers, and the Board looks forward to future submissions.

Note: Many of the photographs, figures, and tables in this volume have been converted from color to grayscale for printing. The electronic files of the papers, posted on the web at **www.TRB.org/TRROnline**, retain the color versions of photographs, figures, and tables as originally submitted for publication.

Measurement Conversion Factors

To convert from the unit in the first column to the unit in the second column, multiply by the factor in the third column.

Customary Unit	*SI Unit*	*Factor*
Length		
inches	millimeters	25.4
inches	centimeters	2.54
feet	meters	0.305
yards	meters	0.914
miles	kilometers	1.61
Area		
square inches	square millimeters	645.1
square feet	square meters	0.093
square yards	square meters	0.836
acres	hectares	0.405
square miles	square kilometers	2.59
Volume		
gallons	liters	3.785
cubic feet	cubic meters	0.028
cubic yards	cubic meters	0.765
Mass		
ounces	grams	28.35
pounds	kilograms	0.454
short tons	megagrams	0.907
Illumination		
footcandles	lux	10.76
footlamberts	candelas per square meter	3.426
Force and Pressure or Stress		
poundforce	newtons	4.45
poundforce per square inch	kilopascals	6.89
Temperature		
To convert Fahrenheit temperature (°F) to Celsius temperature (°C), use the following formula: °C = (°F − 32)/1.8		

SI Unit	*Customary Unit*	*Factor*
Length		
millimeters	inches	0.039
centimeters	inches	0.394
meters	feet	3.281
meters	yards	1.094
kilometers	miles	0.621
Area		
square millimeters	square inches	0.00155
square meters	square feet	10.764
square meters	square yards	1.196
hectares	acres	2.471
square kilometers	square miles	0.386
Volume		
liters	gallons	0.264
cubic meters	cubic feet	35.314
cubic meters	cubic yards	1.308
Mass		
grams	ounces	0.035
kilograms	pounds	2.205
megagrams	short tons	1.102
Illumination		
lux	footcandles	0.093
candelas per square meter	footlamberts	0.292
Force and Pressure or Stress		
newtons	poundforce	0.225
kilopascals	poundforce per square inch	0.145
Temperature		
To convert Celsius temperature (°C) to Fahrenheit temperature (°F), use the following formula: °F = (°C × 1.8) + 32		

Abbreviations Used Without Definitions

AASHO	American Association of State Highway Officials
AASHTO	American Association of State Highway and Transportation Officials
ACRP	Airport Cooperative Research Program
APTA	American Public Transportation Association
ASCE	American Society of Civil Engineers
ASTM	American Society for Testing and Materials (known by abbreviation only)
FAA	Federal Aviation Administration
FHWA	Federal Highway Administration
FMCSA	Federal Motor Carrier Safety Administration
FRA	Federal Railroad Administration
FTA	Federal Transit Administration
IEEE	Institute of Electrical and Electronics Engineers
ISO	International Organization for Standardization
ITE	Institute of Transportation Engineers
NASA	National Aeronautics and Space Administration
NCHRP	National Cooperative Highway Research Program
NHTSA	National Highway Traffic Safety Administration
RITA	Research and Innovative Technology Administration
SAE	Society of Automotive Engineers
SHRP	Strategic Highway Research Program
TCRP	Transit Cooperative Research Program
TRB	Transportation Research Board

Random Coefficient Mixed Logit Models Based on Generalized Antithetic Halton Draws and Double Base Shuffling

Raghuprasad Sidharthan and Karthik K. Srinivasan

Mixed logit models based on quasi-random draws are increasingly being used in discrete choice analysis because of their flexibility. Currently used mixed logit models are too expensive, and their performance degrades with increasing dimensionality. To overcome those shortcomings, two new simple and practical techniques are proposed, namely, quasi Monte Carlo (QMC) with generalized antithetic draws and the double base shuffling method (QMC with generalized antithetic draws and double base shuffling). In a comparison of the performance on probability evaluation, the proposed methods are found to be statistically superior (more accurate and precise) to conventional Halton draws for various dimensions. Results show that proposed methods, unlike conventional Halton draws, are less susceptible to dimensional deterioration even at higher dimensions. Computational experiments with real and synthetic data sets also reveal that the proposed methods are significantly faster for simulated estimation of mixed logit models (at higher dimensions) than other benchmark models [standard Halton, modified Latin hypercube sampling (MLHS), and shuffled Halton draws] to achieve similar accuracy levels. For the real data, the proposed method is 2.1 times faster than conventional QMC for 15 dimensions. The speedup of the proposed methods with synthetic data sets of 15 and 30 dimensions is even greater. The speedup ratio of the proposed methods is 3.3 to 3.4 with respect to conventional Halton draws, and the factor ranges from 2 to 3.2 with respect to MLHS and shuffled Halton draws. Thus, the proposed QMC methods offer promise for the development of richer and more flexible discrete choice models in large dimensional choice contexts.

Discrete choice models have been used to model decision making in econometrics, marketing, finance, and transportation. A commonly used discrete choice model in the transportation planning context is the multinomial logit (MNL) model. The MNL model is popular because of its computational tractability, but assumes independence across alternatives. Consequently, it is inadequate in capturing many realistic features such as random taste variations across observations and flexible correlation structures.

To overcome those shortcomings, recently mixed logit models have been proposed by combining a flexible error-term distribution to the extreme value Type I errors of the MNL model. The estimation of this model is still computationally difficult because it involves the multidimensional integration using Monte Carlo simulation. Although the Monte Carlo estimation is inexpensive for small dimensions ($n < 10$) reported in the literature, there is growing interest in capturing travel decisions of higher dimensions. These arise in the following contexts among others: repeated decisions by the same individual over time and modeling activity and travel decisions of an individual jointly. For these dimensions and contexts, currently used mixed logit models are computationally prohibitively expensive. In the literature quasi Monte Carlo (QMC) methods have been proposed that are faster than the Monte Carlo technique, but their performance deteriorates significantly for large dimensions. Therefore, computationally faster and stable mixed logit models are needed for modeling large dimensional discrete choices.

This paper investigates three related objectives. The first is to propose new methods for generating QMC draws that address the problem of dimensional degradation and are faster and efficient at higher dimensions. In that regard, to overcome those shortcomings two new simple and practical techniques are proposed, namely, generalized antithetic draws (QMCGA) and the double base shuffling method (QMCGADB). The second objective is to investigate the relative performance of the proposed methods for probability evaluation relative to alternative generation techniques [pseudo–Monte Carlo (PMC) and QMC] under varying dimensions (5, 15, and 30). The final objective is to analyze the performance of the proposed methods in regard to likelihood maximization in relation to the standard QMC-based method.

To investigate those objectives, mixed logit models that combine the use of better sampling (generalized antithetic draws) and better shuffling (double base shuffling) techniques are proposed. The accuracy, precision, and computational times of proposed methods for probability evaluation are compared against conventional Halton draws, shuffled Haltons, and modified Latin hypercube sampling (MLHS) draws by using computational experiments. The performance of the best proposed method (QMCGADB) in regard to maximum likelihood estimation (MLE), forecasting ability, and parameter retrieval ability is also investigated by using a large real data set and several synthetic data sets.

The main contribution from this work is the development of improved Halton-based QMC draws that do not deteriorate significantly with increasing dimensions. The proposed methods are faster by a factor of two to three at dimensions as high as 15 to 30, and yield accuracy of probability, parameter estimates, likelihood estimates, and forecasting comparable with conventional Halton draws. These improvements stem from the combined use of multidimensional antithetic variates and shuffling with only two bases of relatively shorter cycle lengths. The benefits at lower dimensions though are

Transportation Engineering Division, Department of Civil Engineering, Indian Institute of Technology, Room 235, Building Sciences Block, Madras, Chennai, 600036, India. Corresponding author: R. Sidharthan, srprasad@gmail.com.

Transportation Research Record: Journal of the Transportation Research Board, No. 2175, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 1–9.
DOI: 10.3141/2175-01

smaller (nearly 35% faster) because the conventional QMC does not display much deterioration here.

The rest of the paper is organized as follows. The next section reviews related work on QMC simulation in relation to the objectives of this study. The methodology of the proposed methods based on the use of general antithetic draws and double base shuffling are presented next. The probability evaluation experiments are then described to compare the performance of the proposed methods and their results, followed by a description of the MLE experiments for the synthetic and real data cases along with a discussion of results. Concluding remarks end the paper.

LITERATURE REVIEW

Mixed logit models try to combine the tractability of MNL with the flexibility of simulated models such as probit. But these are still unlikely to be used for large dimensions because of the computational time involved. Pseudo random numbers were used with mixed logit models initially. The disadvantage of pseudorandom numbers is that they do not give an even coverage of the space, but instead tend to bunch together. Hence, the convergence rate of the pseudorandom numbers is relatively slow.

To overcome the slow convergence, Train (*1*) and Bhat (*2*) proposed the use of Halton draws for lower dimensions (1–5), which use nonrandom but more uniformly distributed sequences. This property makes the numerical integration converge faster [Train (*1*) and Bhat (*2*)]. Although the Halton sequences performed well at low dimensions (<6), performance deteriorated at higher dimensions. At these dimensions, these draws become highly correlated as a result of their cyclical nature and give poor coverage [Hess et al. (*3*)]. These limitations of standard Halton draws with increasing dimensions are widely acknowledged. Because of the limitations, several researchers have proposed improved QMC methods of mixed logit models. Broadly, the improvements may be categorized as (*a*) improvements to standard Halton draws and (*b*) use of non-Halton QMC draws.

Among Halton-based approaches, two types of improvements have been proposed to standard Halton draws. The first line of studies is based on scrambled Halton draws. Along that line, Bhat compared the performance of scrambled Halton and Halton draws in the context of a mixed probit model with three alternatives, 10 dimensions, and 22 parameters by using synthetic data (*4*). He reported that scrambled Haltons were better than standard Halton draws, which in turn, outperformed PMC draws. Sivakumar et al. investigated the performance of standard Halton, scrambled Halton, standard Faure, and scrambled Faure (*5*). The computational study was conducted with two synthetic data sets with five (up to 625 draws) and 10 dimensions (100 draws). They found that the scrambled Faure performed better than standard and scrambled Halton draws.

The second Halton-based improvement technique is based on the concept of shuffled Halton draws. Here, to overcome the problem of correlations at higher dimensions, Halton sequences are generated from a single prime base and are randomly shuffled to obtain draws corresponding to different dimensions. Along that line Hess and Polak demonstrated that the random shuffling technique can help in reducing the problem of correlations and incomplete cycles in higher dimensions observed for standard Halton draws (*6*). They noted that shuffled Halton draws offer significant promise of computational time saving, better coverage, and correlation reduction compared with scrambled Halton draws. Hess et al. compared the performance of standard, scrambled, and shuffled Halton sequences in mixed logit models involving synthetic and real data [five-dimensional (5-D)] (*7*). They observed that shuffled Halton sequences are more reliable than scrambled Halton. Wang and Kockelman compared the performance of various types of shuffling relative to scrambled Haltons and PMC for a two-dimensional, three-alternative spatial choice context with correlated observations (*8*). Their results showed that with increasing draws (200 or more), the relative performance of shuffled, scrambled, and standard Halton draws were relatively close and one of the shuffled versions led to a lower bias than scrambled draws.

A number of QMC alternatives to Halton draws have also been investigated. Hess et al. proposed MLHS (*3*). The MLHS method is practical and simple to implement, and it performed better than both scrambled and shuffled Halton draws in a relatively large dimensional context (16 dimensions). Garrido studied the relative performance of Sobol draws with respect to Halton and PMC draws (*9*). By using five- and 10-dimensional synthetic data sets, it was found that 150 Sobol draws were more accurate than 200 Halton draws. Sandor and Train compare the performance of four kinds of (t,m,s)-nets, standard Haltons, and Haltons with random start by using a 5-D mixed logit context (*10*). They reported that two of the (t,m,s)-nets were better than Halton draws. Bastin et al. compared the performance of PMC, Sobol, and scrambled Sobol and found the latter methods to be more accurate than PMC for a 5-D case (*11*).

From these studies, it is clear that shuffling and scrambling help in reducing the correlation present across dimensions in conventional Halton draws. Further, shuffled Haltons, scrambled Faure, and MLHS techniques outperform conventional and scrambled Halton draws for lower dimensions (<10) and few parameters (<25). However, there is limited insight on the performance of alternative QMC methods in higher dimensional contexts with relatively large numbers of parameters. Furthermore, most of these studies evaluate the performance of simulated maximum likelihood estimates and optimized log likelihood (LL). Few studies also analyze the probability evaluation performance of alternative QMC methods [e.g., Train (*1*) and Bhat (*2, 4*)], but only with regard to optimal parameters and chosen alternative. Therefore, an understanding of the performance over different probability ranges and parameter ranges is lacking. Consequently, the role of the number of draws and asymptotic performance of QMC methods on probability evaluation performance is not well understood.

PROPOSED QMC METHODS: QMC WITH GENERALIZED ANTITHETIC DRAWS AND DOUBLE BASE SHUFFLING

In standard Halton draws, the draws for different dimensions are generated from different prime bases. This leads to correlation across higher dimensions and causes poor coverage resulting from incomplete cycles. Each Halton draw vector is intended to give a vector of independent and uniformly distributed variables (U), which are converted into a corresponding vector of standard normal random variable (Z) by normal inversion or other procedures.

QMC with Generalized Antithetic Halton Draws

The principle of the antithetic variate technique is to induce a negative correlation between estimates obtained from different draws, such that the variance of the overall estimate is reduced. Although antithetic variates have been used for probit models, their application to the mixed logit model is not significant.

A new antithetic scheme referred to as the generalized antithetic draws is proposed here to yield more normal vectors from the same draw. From the vector z = (z1r, z2r, z3r, z4r, z5r), five other antithetic vectors with the same distribution can be obtained by permuting the terms element by element as za = (−z1r, z2r, z3r, z4r, z5r), zb = (z1r, −z2r, z3r, z4r, z5r), zc = (z1r, z2r, −z3r, z4r, z5r), zd = (z1r, z2r, z3r, −z4r, z5r), ze = (z1r, z2r, z3r, z4r, −z5r). Thus, for each realization r, six probability estimates can be obtained corresponding to each of the normal vectors and averaged for this 5-D example. A randomized variant of the generalized antithetic scheme above is proposed and applied in this study. First, the vector Z is generated as in standard Halton draws as noted above. Next, $m-1$ additional antithetic normal vectors are generated randomly. For each dimension, randomly generate a multiplier of −1 or +1 and apply these multipliers to the vector Z to obtain a new vector Z′. This process is repeated $m-1$ times. This procedure ensures that the distribution of the antithetic vectors is also identical to that of the original vector. Thus, for each base draw vector, $m-1$ antithetic and 1 base probability values are estimated and averaged. From computational investigations, choice of a randomized subset of $m=5$ antithetic vectors was found to be sufficiently accurate for dimensions up to 30.

The rationale for the generalized antithetic draws is threefold. First, with a given number (k) of Halton draws, mk normal draws can be obtained and used for probability evaluation leading to computational time savings in generation (which is fivefold for $m=5$). Second, the presence of negative multipliers for some dimensions will lead to negative correlation between the antithetic normal draws and, hence, reduce the variance of probability estimates over draws. Finally, the dimensions that form the antithetic variables are randomized across draws leading to better coverage over multidimensional space.

The expression for the mixed logit probability for the alternative specific taste variation case (which will be the focus in this study) is given below (Equations 1 and 2):

$$P_i = \int \frac{\exp\left(\sum_{k=1}^{T} \beta_{k,i} x_{k,i}\right)}{\sum_{j=1}^{N} \exp\left(\sum_{k=1}^{T} \beta_{k,j} x_{k,j}\right)} f(\beta) d\beta \quad (1)$$

where

$\beta_{k,i}$ = linear utility coefficient of the attribute k for alternative i,
$x_{k,i}$ = kth attribute value for alternative i,
$f(\beta)$ = distribution of β,
T = number of attributes, and
N = number of alternatives.

For the special case where $f(\beta)$ is normally distributed with $N(b, \sigma)$ the expression for probability becomes

$$P_i = \int \frac{\exp\left(\sum_{k=1}^{T} b_{k,i} x_{k,i} + \sum_{k=1}^{T} \sigma_{k,i} \eta_{k,i} x_{k,i}\right)}{\sum_{j=1}^{N} \exp\left(\sum_{k=1}^{T} b_{k,j} x_{k,j} + \sum_{k=1}^{T} \sigma_{k,j} \eta_{k,j} x_{k,j}\right)} f(\eta) d\eta \quad (2)$$

where

$\beta_{k,i} = b_{k,i} + \sigma_{k,i}\eta_{k,i}$,
$b_{k,i}$ = mean of the distribution of coefficient $\beta_{k,i}$,
$\sigma_{k,i}$ = standard deviation of coefficient $\beta_{k,i}$, and
$\eta_{k,i}$ = standard normal distribution.

QMCGA with Double Base Shuffling

As mentioned earlier, the drawback with Halton sequences generated from higher primes is that as the cycle length (equal to the value of the prime) becomes high it gives rise to correlation with other high prime sequences. The second problem is that of incomplete cycles. When the cycle length increases, the cycles in many of the dimensions remain incomplete. This changes the mean of the expected value and creates bias in the estimates.

To address that problem, the double base shuffling scheme is proposed. The proposed QMCGADB is different from standard Halton draws in regard to generation of uniform draws as noted below. But once these are drawn, the generalized antithetic procedure discussed above is applied to convert them into normal draw vectors. Unlike conventional Halton draws, in which the draw for each dimension comes from different prime number bases, in double base shuffling, all numbers are generated from only two prime bases (Base 2 and 3).

The principle behind the double base shuffling is to take advantage of the progressively uniform coverage that Halton sequences provide, while not compromising the quality as dimensionality increases. Because the prime bases used are only 2 and 3, the combined cycle length (6) is extremely short and is quickly completed. So even with fewer draws (say, 100), the effect of incomplete cycles is significantly reduced. The choice of lower numbered bases along with the shuffling reduces the correlation that may exist between any two of the dimensions. Figure 1 shows the first 300 draws obtained for the dimensions 29 and 30 by using standard Haltons, shuffled Halton, scrambled Sobol, scrambled Faure, MLHS, and double base shuffling (DBSH). The incomplete coverage of the space can clearly be seen for the standard Haltons, whereas DBSH covers the space more evenly. A similar coverage is also noted for shuffled Haltons and MLHS. The QMC estimation procedure based on combining DBSH and generalized antithetic draws is referred to as QMCGADB. The main advantage of QMCGADB is that it combines the reduced generation time of QMCGA and improved performance of the DBSH method at higher dimensions.

PROBABILITY EVALUATION EXPERIMENTS

Computational experiments are performed to evaluate the proposed methods at the level of probability estimation because its accuracy is critical to the maximum likelihood estimation procedure. Further, performance measures at this stage are not influenced by the performance of the optimization procedure.

Experiment Description

In these experiments the performance of the following methods is evaluated: PMC, QMC, QMCGA, and QMCGADB. This evaluation is carried out for three different dimensions of integration: 5, 15, and 30. For each dimension, the probability evaluation is carried out for several draws (200, 400, 800, 1,600, and 3,200). The performance measures used for comparison include accuracy [root mean square error (RMSE) percentage] and precision (standard deviation) of probability estimates and computational time. The true probability used for benchmarking is obtained by using PMC with 640,000 draws.

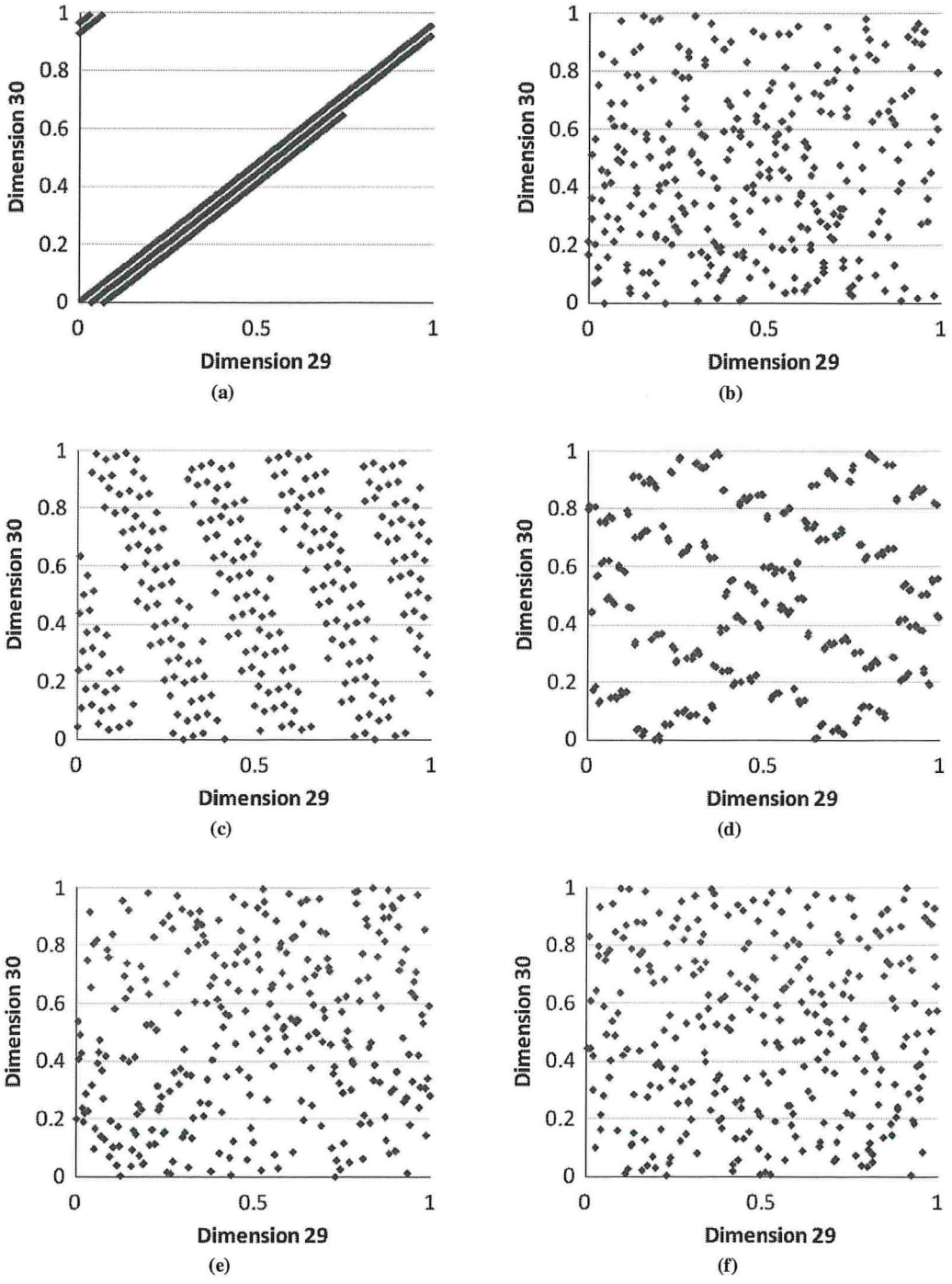


FIGURE 1 Scatter plot of first 300 points from Dimensions 29 and 30 obtained from different generation strategies: (*a*) Halton, (*b*) shuffled Halton, (*c*) scrambled Faure, (*d*) scrambled Sobol, (*e*) MLHS, and (*f*) DBSH.

The probability evaluation experiment for each dimension is conducted by using 50 observations generated randomly. Each observation was generated to allow a wide range of utility values for the alternatives and different taste variation parameters. For each observation, 25 replications of the probability estimates were obtained by using each of the methods. For the PMC, the replications were obtained by using different seed values, whereas for the QMC methods, successive Halton sequences were used to obtain the replications.

For each observation, by using the 25 replications, the average error and the variance of probability estimates are computed. The average error is a measure of bias in probability evaluation, and the variance is a measure of simulation noise. The overall inaccuracy (RMS %) is obtained through the RMS of the average probability errors of all 50 observations. The average of the variances across observations is taken as the measure of precision. Another auxiliary performance measure, namely, the speedup factor, is also used in

comparison. The speedup factor is defined as the ratio of computational time of the baseline method to the given method for a given level of accuracy (RMSE).

The convergence and small sample behavior for each method is analyzed by studying the rate at which the accuracy increases (RMSE R decreases) with increasing draws (N). In particular, this is investigated by calibrating the equation $R = aN^b$. The coefficient b represents the asymptotic convergence rate, and the constant a indicates the convergence speed at a lower number of draws. In other words, a smaller constant a and larger exponent b (absolute value) signify faster convergence. These results are discussed in the following sections.

Results for 5-D Case

The 5-D case results (Figure 2*a*) confirm findings from other lower dimensional studies that the probability estimate using QMCGA and

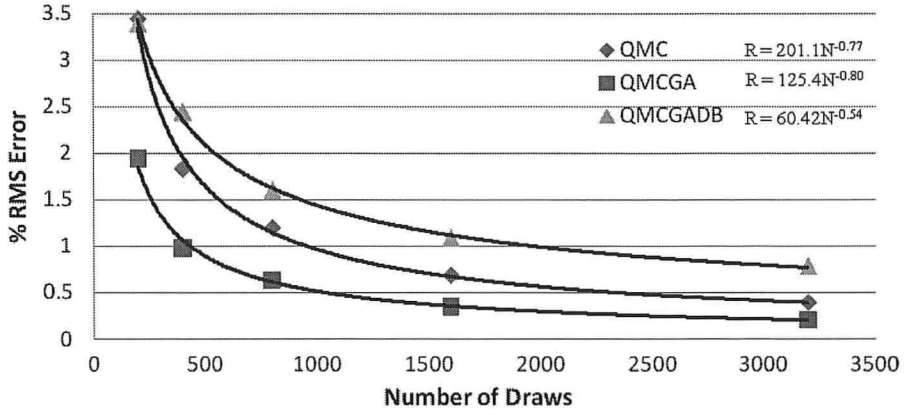


FIGURE 2 Asymptotic performance of RMS error with increase in number of draws for (*a*) 5-D, (*b*) 15-D, and (*c*) 30-D.

TABLE 1 Relative Accuracy and Precision of Proposed Methods for Varying Dimensions and Draws

	5-D			15-D			30-D		
Method	100	200	400	100	200	400	100	200	400
RMS % (difference between estimated probability and "true" probability)									
PMC	16.1918	11.4020	8.0292	9.2828	6.5532	4.6262	7.4376	5.2853	3.7558
QMC	5.7150	3.3439	1.9566	6.5325	3.4861	1.8603	4.1041	2.7168	1.7985
QMCGA	**3.2065**	**1.8469**	**1.0638**	3.3462	1.8287	0.9993	2.8103	1.7588	1.1008
QMCGADB	5.0020	3.4378	2.3627	**1.8511**	**1.2831**	**0.8894**	**1.3985**	**0.9303**	**0.6188**
Precision (SD of prob.)									
PMC	0.1600	0.1125	0.0791	0.0910	0.0642	0.0452	0.0721	0.0513	0.0365
QMC	0.0574	0.0334	0.0194	0.0680	0.0354	0.0184	0.0405	0.0268	0.0177
QMCGA	**0.0331**	**0.0186**	**0.0104**	0.0369	0.0189	0.0097	0.0285	0.0175	0.0107
QMCGADB	0.0487	0.0335	0.0230	**0.0185**	**0.0126**	**0.0086**	**0.0145**	**0.0093**	0.0060

NOTE: SD = standard deviation. Values in bold indicate the best-performing method at the specified dimension level and number of draws.

QMC are more accurate than PMC for a given number of draws. The asymptotic convergence rate of QMCGA ($N = -0.80$) is better than QMC ($N = -0.77$), whereas QMCGADB ($N = -0.541$) and PMC ($N = -0.51$) were much slower. In contrast to the asymptotic rate, the initial convergence rate is a more important factor when fewer draws are used or lower accuracy levels are specified. This factor is the best for QMCGADB (60.42), followed by QMCGA (125.4) and PMC (166) compared with QMC (201). QMCGA is also more accurate and precise than other methods (see Table 1). The accuracy of QMCGA is 78% to 84% higher and precision is 73% to 86% greater than QMC.

For a 4% RMSE level, the QMCGA has a speedup factor of 1.5 relative to QMC, implying that QMC takes 50% more time than QMCGA. For a higher level of accuracy (0.1% RMSE), this speedup increases slightly to 1.7. However, for the 5-D case, QMCGADB has a speedup factor of less than 1 at all accuracy levels. Thus, QMCGA with shuffling is neither necessary nor productive at lower dimensions because it removes the even coverage. So for five dimensions, QMCGA is the preferred method.

Results for 15-D Case

The degradation of the standard Halton draws is observed with increasing dimensions (Figure 2*b*). The asymptotic convergence rate of QMCGA is –0.87, which is almost close to that of QMC (–0.91), but the initial convergence factor (constant) is much smaller (by a factor of more than 2) for QMCGA, giving it the edge. Compared with this, QMCGADB has a lower asymptotic convergence rate of only about –0.53, but that is compensated for by a much faster initial convergence rate (20 times better than QMC).

Both QMCGA and QMCGADB are more accurate (by 1.9 to 3.5 times) and precise (by 1.8 to 3.67 times) than QMC for a given number of draws (Table 1). For RMS error levels less than 1% (higher accuracies) QMCGA has a higher speedup factor relative to both QMCGADB and QMC, and at 0.1% error levels it is as high as 1.5, whereas at RMS error levels higher than 1%, QMCGADB has the highest speedup factor, reaching a maximum of 4.5 at a 4% error level. Thus, QMCGA is the preferred method for RMS of <1%, whereas QMCGADB is the preferred method for RMS errors >1%. With fewer draws, QMCGADB provides a better performance because it has more completed than uncompleted cycles, unlike QMCGA, and the role of the initial convergence factor dominates over the asymptotic rate coefficient.

Results for 30-D Case

At all accuracy levels, QMCGADB is superior to the other methods in both accuracy and precision for the thirty-dimensional (30-D) case (Figure 2*c*). The RMS error for QMCGADB is smaller by a factor of 2.9 to 2.93 relative to QMC (100 to 400 draws) (Table 1). The accuracy improvement for QMCGA (relative to standard Haltons) ranges between 1.52 and 1.7. The precision improvement ratios (relative to standard Haltons) are 2.79 to 2.97 for QMCGADB and 1.36 to 1.56 for QMCGA. The speedup factor for QMCGADB (relative to QMC) ranges from 3 at the 4% RMS error level to about 3.2 at the 0.1% RMS error level. The same for QMCGA varies from about 1.2 at 4% RMS error levels to about 2.1 at 0.1% error levels. Thus, QMCGA, although not as good as QMCGADB, also mitigates the dimensional degradation observed in standard QMC. For a given accuracy level, QMCGADB is faster than other methods, followed by QMCGA.

EXPERIMENTS BASED ON MAXIMUM LIKELIHOOD ESTIMATION

This section investigates the performance of the proposed methods in maximum likelihood estimation (MLE) and parameter retrieval accuracy relative to conventional QMC draws. As noted earlier, QMCGA is the recommended method for five and 15 dimensions and QMCGADB for 30 dimensions. To enable comparison, the draws for different methods are chosen such that the accuracy of likelihood computations is nearly the same at the true parameters. This ensures that the MLE results obtained from the different models are equivalent and that their computation times can be directly compared as a measure of estimation efficiency. The trust region method with analytically coded gradients is used for MLE, and the hessian is approximated by using the outer product of gradients.

This investigation is performed with synthetically generated data sets, as well as a real-world data set. In the synthetic data experiments, three different synthetic data sets are constructed corresponding to

5, 15, and 30 dimensions. For each dimension, the data consists of 10 sets of 2,000 observations obtained from a given utility model with random taste variation. The key performance indicators are the average maximum LL, the inaccuracy in the estimated parameters (average absolute mean percentage error across parameters), and the time taken. In addition, the maximum LL is compared with the true-parameter LL. Further, the forecasting performance on a holdout data set is also analyzed. In the real-world data set, the true parameters are unknown. Hence, the LL value, computation time, and forecasting performance are compared. The parameter estimates from the two methods are also studied.

Synthetic Data Experiments and Results

The 5-D case corresponds to a five-alternative choice context with one explanatory variable. This variable is assumed to have a generic mean and alternative specific random taste variation leading to a 5-D Monte Carlo integration. For the 15-D case, also, a five-alternative context is considered, but with two generic variables, 10 alternative specific random taste parameters, and a heteroskedastic error structure (adding to five more dimensions of integration). The 30-D case is constructed in a 10-alternative choice context with two generic variables, 20 alternative specific taste parameters, and 10 heteroskedastic error terms.

The distribution for the data variables (attributes) and the parameters (beta, sigma) were chosen to ensure their statistical significance in the model. For the 15-D case, the coefficients (betas) for the two generic variables were chosen as −1.1 and −2.65. The corresponding variables were generated randomly from a uniform distribution with a mean of 4.75 and a spread of 7.5 for the first variable and a mean of 5.5 and a spread of 10 for the second variable. The alternative specific constant for the first alternative was fixed to zero, and others were generated from a uniform distribution with a mean of 2.3 and a spread of 3. The alternative specific taste variation parameters, sigma (standard deviation), for the first variable were chosen randomly between 0.02 and 0.60 for each of the five alternatives; for the second variable it was selected randomly between 0.15 and 1.15.

The comparison of the key performance indicators is shown in Table 2. In addition to QMC and QMCGADB, two other benchmark models are considered for higher dimensions (MLHS and shuffled Halton draws). The maximum LL values were comparable between the proposed methods and QMC with a significant reduction in computation time. For the 5-D case, QMC is 35% slower than QMCGA. But for 15- and 30-D cases, the QMCGADB is much faster (the speedup factor is 3.35 to 3.42 compared with QMC). For the 15-D case, the proposed method is faster, by factors of 3.07 and 3.24, than MLHS and shuffled Halton draws, respectively. In the 30-D case, the speedup ratios of the proposed model relative to shuffled Halton and MLHS draws were 1.99 and 2.54, respectively. In contrast to the expectation that larger savings may be found in 30-D compared with 15-D because of the greater deterioration of standard Halton draws, it was found that the speedup factor is roughly the same. Closer investigation revealed that the Halton generation component savings were indeed larger for 30-D. However, this saving is offset partly by an increase in the fraction of computation time associated with gradient computation and optimization at larger dimensions.

The forecasting ability of the estimated parameters is tested by calculating the LL of the 10th data set (by using the average parameters obtained from the first nine data sets). For the 5-D case, the QMC results give a forecast LL of −2437.28 compared with −2437.34 by the QMCGADB method. The true parameter LL for this case is −2436.34, indicating a good forecasting ability of the estimated parameters. For the 15-D case the forecast LL for the QMC and QMCGADB are −1505.61 and −1504.94, respectively, and the true parameter LL is −1498.34. For the 30-D case the forecast LL is −2109.2 and −2103.1 for QMC and QMCGADB. The true parameter LL was −2084.9.

The accuracy of the parameter estimates (absolute mean percentage error) is obtained by comparing against the true parameters used in the

TABLE 2 Comparison of MLE Performance of Proposed Method Versus Alternative Benchmark Models for 15-D and 30-D (Synthetic Data) and 15-D (Real Data)

Method	Estimation LL	Computational Time (min)	Proposed Model Speedup Relative to Other Models
15-D Synthetic			
MLHS	−1,526.3	7.97	3.07
Proposed (QMCGADB)	−1,528.9	2.6	1
QMC standard	−1,531.5	8.9	3.42
Shuffled Haltons	−1,533.5	8.43	3.24
30-D Synthetic			
MLHS	−2,040.2	12.45	2.54
Proposed (QMCGADB)	−2,047.4	4.9	1
QMC standard	−2,050.5	16.4	3.35
Shuffled Haltons	−2,042.4	9.75	1.99
15-D Real Data			
MLHS	−3,532.1	70.89	1.52
Proposed (QMCGADB)	−3,527.4	46.75	1
QMC standard	−3,522.2	100.08	2.14
Shuffled Haltons	−3,526.0	80.72	1.73

generation. For all dimensions, the accuracy of QMC and QMCGADB are comparable. They are also nearly equal to those of other benchmark models. For the 5-D case, the bias in the estimated parameters is very low and the average error across the parameters is only 4.7% for QMC and 4.8% for QMCGADB. But for the higher dimensions (15-D and 30-D) the parameters are more inaccurate. For 15-D the average error across parameters is 158% for QMC and 11% for QMCGADB; the average error across parameters for the 30-D case is 111% for QMC and 102% for QMCGADB. At these higher dimensions, the betas are the least biased grouping with about 25% to 50% error, whereas the taste variation parameters are the most biased group with biases in the range of 130% to 250%. These biases are present despite the optimized likelihood being as good as the true parameter likelihood, suggesting the possibility of multiple optima at higher dimensions.

MLE Comparison with Real Data

The route choice data collected by using the stated choice experiments [Srinivasan and Mahmassani (*12*)] were used to compare the standard QMC method and QMCGADB method for maximum likelihood estimation. In this choice context, users chose one from among three alternative routes—a freeway, a major arterial, and a minor arterial—on the basis of different types of information and the nature of the information provided. Users were supplied with travel times and visual congestion on the different routes.

Three experimental factors (with different levels) concerning advanced traveler information system information quality and credibility were examined in this study. These include type of information (descriptive and prescriptive), credibility of information (feedback provided on recommended path, best path, and no feedback), and quality of information (classified on the basis of prevailing, predicted, random, perturbed, and differential levels).

The data set consisted of a total of 6,396 observations of which 75% were used for calibration. The final model has 12 generic beta attributes and 15 alternative specific random taste variation parameters leading to a 15-D mixed logit model. The LLs of QMC and QMCGADB are comparable at −3522.16 and −3527.37, respectively. The proposed QMCGADB was found to be nearly twice as fast (47 min compared with 100 min) (Table 2). Furthermore, the proposed model was also considerably faster than other benchmark models. The speedup factors relative to MLHS and shuffled Halton draws were 1.52 and 1.73, respectively. The forecast LLs are also close (QMC LL = −1208.7, QMCGADB LL = −1211.8) and better than that of the MNL model (LL = −1285.1). The estimated parameters and the *t*-statistics are shown in Table 3, for both QMC and QMCGADB, and are reasonably close in most cases. The signs and (in)significance match in the two models at the 10% level.

The empirical findings are as follows. As expected, travel time and congestion have negative coefficients and are highly significant. The preference heterogeneity for travel time is much higher for freeway (Alternative 1) compared with major and minor arterial roads. The coefficient of prescriptive information is negative, implying greater choice probability for freeways than for arterials when prescriptive information is given. The results also suggest that providing feedback on the recommended and best routes' attributes decreases the attractiveness of arterials compared with freeways. The estimates for predicted and prevailing levels are negative, implying that freeways are preferred under these information types. The last four variables

TABLE 3 Estimated Parameters and *t*-Statistic for Real Data with QMC Method and QMCGADB Method

	QMC				QMCGADB			
Parameter	Beta	*t*-Stat.	Sigma	*t*-Stat.	Beta	*t*-Stat.	Sigma	*t*-Stat.
Travel time	−0.146	−4.641	0.082(r1) 0.203(r2) 0.202(r3)	4.310 4.462 4.520	−0.138	−4.207	0.076(r1) 0.179(r2) 0.193(r3)	3.349 4.074 4.129
Congestion	−6.418	−4.251	3.133(r1) 5.076(r2) 2.429(r3)	3.871 4.341 1.365	−6.104	−4.116	3.115(r1) 5.103(r2) 2.371(r3)	3.827 4.174 1.466
Prescriptive	−0.788	−3.236			−0.769	−3.219		
Recommended	*−0.108*	*−0.480*			*−0.123*	*−0.571*		
Best	−0.253	−1.193			−0.245	−1.197		
Predicted	−0.720	−2.600			−0.636	−2.405		
Prevailing	−0.729	−3.034			−0.671	−2.913		
Differential	*−0.169*	*−0.448*			*−0.185*	*−0.510*		
Cong * best	*−0.382*	*−0.696*	0.165(r1) 0.612(r2) 5.611(r3)	0.217 0.244 3.275	*−0.840*	*−1.317*	1.028(r1) 1.489(r2) 5.493(r3)	1.285 0.558 3.309
Cong * recommended	−1.010	−1.801			−0.847	−1.528		
Cong * prevailing	−5.370	−4.694	3.021(r1) 0.607(r2) 0.563(r3)	3.018 0.278 0.121	−5.750	−4.311	3.245(r1) 0.057(r2) 0.007(r3)	4.051 0.014 0.001
Cong * predicted	−8.116	−3.519	5.238(r1) 7.273(r2) 7.264(r3)	3.914 3.234 2.974	−6.333	−2.958	3.985(r1) 5.517(r2) 5.932(r3)	2.679 2.451 2.126

NOTE: The standard deviations for r1 = freeway, r2 = major arterial, and r3 = minor arterials are also shown. Coefficients in italics are insignificant at 10% level.

are interaction variables between visual congestion information and information factors (type, quality, and feedback) above. The negative signs of these variables indicate greater sensitivity to visual congestion under predicted information.

CONCLUSIONS

Two new methods, namely, QMCGA draws and QMCGADB shuffling based on Halton draws, are proposed in this paper. Generalized antithetic variates improve the accuracy and precision of the probability estimates obtained from each Halton draw, and double base shuffling reduces the cycle length to six, irrespective of the dimensionality of the problem. Thus the proposed method overcomes the problems of incomplete cycle length and correlation between dimensions at higher dimensions encountered in standard Halton draws. These methods are effective in reducing the variance of probability estimates and the correlation across higher dimensions noted with conventional Halton draws. The proposed methods are found to be statistically superior (more accurate and precise) than conventional Halton draws for various dimensions. Further, the proposed methods are less susceptible to dimensional deterioration even at higher dimensions, unlike conventional Haltons.

The different methods of Monte Carlo evaluation are compared at the level of probability calculation for various dimensions (5-D, 15-D, and 30-D). The salient findings from the probability evaluation experiments are as follows: The speedup factor of the proposed methods compared with QMC increases from 1.5 at 5-D to as high as 3.5 at 30-D, illustrating the ability to address dimensional degradation. The proposed methods exhibit less dimensional deterioration from 15-D to 30-D. The deterioration in the asymptotic rates in going from 15 to 30 dimensions is −0.31(the rate decreases from −0.91 to −0.60), −0.18, and +0.06, respectively, for QMC, QMCGA, and QMCGADB. The initial convergence constant, which governs the small sample performance, is very small (better) for the proposed methods, resulting in fewer draws for the same accuracy. For small dimensions (5-D), QMCGA is the preferred method for up to 3,200 draws; QMCGADB is the preferred generation method for higher (15-D and 30-D) dimensions.

In regard to MLE of mixed logit models, results reveal that the proposed methods substantially outperform improved Halton and non-Halton techniques in synthetic and real data sets. For the real data with 15 dimensions, the proposed model is 2.14 times faster than standard Halton draws for equivalent LL accuracy and parameter estimates. For the synthetic data with 30 dimensions, the QMCGADB is faster than standard Halton draws by a factor of 3.35. It is also faster than MLHS (by 2.54 times) and shuffled Halton (by 1.99 times). For the synthetic data sets, the parameters are recovered well at 5-D, but the absolute mean percentage errors in parameters are significantly higher at 15- and 30-D for all methods (including QMC and QMCGADB), even though likelihood and forecasting performance are satisfactory.

From a practical perspective, these improvements offer the promise for quicker and more accurate estimation of large dimensional mixed logit models. This study illustrates that by using the proposed techniques, a 30-D mixed logit model with 2,000 observations and 39 parameters can be estimated in about 5 min of computation time on a personal computer. Thus, the proposed methods are promising for estimating mixed logit models in large dimensional choice contexts, such as destination choice, activity, and tour sequencing as well as panel data. Potential future research could be the study of the applicability of QMCGADB to random taste variation for nonnormal distributions (such as log-normal) and cases in which different parameters possess different distributions.

ACKNOWLEDGMENTS

This research is supported in part by funding from the Centre of Excellence in Urban Transport at the Indian Institute of Technology, Madras, which is funded by the Ministry of Urban Development, government of India. This support is gratefully acknowledged.

REFERENCES

1. Train, K. E. *Halton Sequences for Mixed Logit.* Technical paper. Department of Economics, University of California, Berkeley, 1999.
2. Bhat, C. R. Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model. *Transportation Research Part B,* Vol. 35, No. 7, 2001, pp. 677–693.
3. Hess, S., K. E. Train, and J. W. Polak. On the Use of a Modified Latin Hypercube Sampling (MLHS) Method in the Estimation of a Mixed Logit Model for Vehicle Choice. *Transportation Research Part B,* Vol. 40, No. 2, 2005, pp 147–163.
4. Bhat, C. R. Simulation Estimation of Mixed Discrete Choice Models Using Randomized and Scrambled Halton Sequences. *Transportation Research Part B,* Vol. 37, No. 9, 2003, pp. 837–855.
5. Sivakumar, A., C. R. Bhat, and G. Ökten. Simulation Estimation of Mixed Discrete Choice Models with the Use of Randomized Quasi-Monte Carlo Sequences: A Comparative Study. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1921,* Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 112–122.
6. Hess, S., and J. W. Polak. An Alternative Method to the Scrambled Halton Sequence for Removing Correlation Between Standard Halton Sequences in High Dimensions. Presented at European Regional Science Conference, Jyväskylä, Finland, 2003.
7. Hess, S., J. W. Polak, and A. J. Daly. On the Performance of Shuffled Halton Sequences in the Estimation of Discrete Choice Models. Presented at European Transport Conference, Strasbourg, France, 2003.
8. Wang, X., and K. M. Kockelman. Maximum Simulated Likelihood Estimation with Spatially Correlated Observations: A Comparison of Simulation Techniques. Presented at 53rd North American Regional Science Association International Conference, Toronto, Ontario, Canada, 2005.
9. Garrido, R. A. Estimation Performance of Low Discrepancy Sequences in Stated Preferences. Presented at 10th International Conference on Travel Behaviour Research, Lucerne, Switzerland, Aug. 2003.
10. Sandor, Z., and K. Train. Quasi-Random Simulation of Discrete Choice Models. *Transportation Research Part B,* Vol. 38, No. 4, 2004, pp. 313–327.
11. Bastin, F., C. Cirillo, and P. L. Toint. Application of an Adaptive Monte Carlo Algorithm to Mixed Logit Estimation. *Transportation Research Part B,* Vol. 40, No. 7, 2006, pp. 577–593.
12. Srinivasan, K., K. Mahmassani, and S. Hani. Analyzing Heterogeneity and Unobserved Structural Effects in Route-Switching Behavior Under ATIS: A Dynamic Kernel Logit Formulation. *Transportation Research Part B,* Vol. 37, No. 9, 2003, pp. 793–814.

The Transportation Demand Forecasting Committee peer-reviewed this paper.

Policy Evaluation in Multiagent Transport Simulations

Dominik Grether, Benjamin Kickhöfer, and Kai Nagel

In democratically organized societies, the implementation of measures with regressive effects on welfare distribution tends to be complicated because of low public acceptance. The microscopic multiagent simulation approach presented in this paper can help to design better solutions in such situations. Income can be included in utility calculations for a better understanding of problems linked to acceptability. This paper shows how the approach can be used in policy evaluation when income is included in user preferences. With the MATSim framework, the implementation is tested in a simple scenario. Furthermore the approach works in a large-scale, real-world example. On the basis of a hypothetical price and speed increase of public transit, effects on the welfare distribution of the population are discussed. This approach, in contrast with applied economic policy analysis, allows choice modeling and economic evaluation to be realized consistently.

Policy measures in transportation planning aim at improving the system as a whole. Changes to the system that result in an unequal distribution of the overall welfare gain are, however, hard to implement in democratically organized societies. Studies indicate that, for example, tolls tend to be regressive if no redistribution scheme is considered at the same time and may so increase the inequality in welfare distribution (*1*). An option to reach broader public acceptance for such policies may be to include the redistribution of total gains into the scheme. Hence, methods and tools are needed that simulate welfare changes due to policies on a highly granulated level, that is, considering each individual of the society. With such tools, policy makers are able to consider the effects of different proposed measures on the welfare distribution. In addition, it is possible to estimate the support level in the society and, if necessary, to evaluate alternatives for further discussion.

Traditional transport planning tools using the four-step process combined with standard economic appraisal methods [e.g., Pearce and Nash (*2*)] are not able to provide such analysis. To bridge this gap, multiagent microsimulations can be used. Large-scale multiagent traffic simulations are capable of simulating complete day plans of several million individuals (agents) (*3*). In contrast to traditional models, all attributes attached to the synthetic travelers are kept during the simulation process, thus enabling highly granulated analysis (*4*). Being aware of all attributes makes it possible to attach to every traveler an individual utility function that is used to maximize the individual return of travel choices during the simulation process. Another advantage of the multiagent simulation technique is the connection of travelers' choices along the time axis when time-dependent policies are simulated (*5*).

In the context of policy evaluation, simulation results can immediately be used to identify winners and losers because the utility scores of the individual agents are kept and can be compared between scenarios agent by agent. They can also be aggregated in arbitrary ways on the basis of any available demographic attributes, including spatial information of high resolution. Welfare computations, if desired, can be done on top of that, without having to resort to indirect measures such as link travel times or interzonal impedances. The usual problems when monetarizing the individual utility still apply (*6*), but at least one main issue in applied economic analysis is addressed: with multiagent approaches, choice modeling and economic evaluation are implemented in a consistent framework, similar to efforts to base such analysis directly on discrete choice models (*7*).

This paper shows how multiagent approaches can be used in policy evaluation adding an individual income attribute to each agent so that personalized utilities are considered. It is shown how one benefits doing so when issues linked with public acceptance are considered. Implications of the simulation model are explained, and the measurement of welfare effects resulting from policy measures is highlighted. In a simple scenario, income is the only varying attribute between agents. A real-world scenario, however, includes varying trip distances and day plans so that demographic attributes of each agent are strongly personalized.

The paper is organized as follows. First, the simulation structure is introduced. Then, the income-dependent utility function used in this paper is presented. Afterward, in a simple scenario, the correctness of implementation and the plausibility of results are tested. Subsequently, a realistic simulation of regular workday traffic in the metropolitan area of Zurich, Switzerland, is performed, including the effects of a public transit price and speed increase. Afterward, welfare changes across the income range and open issues are discussed. The paper ends with concluding remarks.

SIMULATION STRUCTURE

The following describes the structure of the simulation that is used. It is the standard structure of MATSim (www.matsim.org), as described in much of the literature (*8, 9*). Readers familiar with the MATSim approach can skip this section.

Transport Systems Planning and Transport Telematics, Berlin Institute of Technology, D-10587 Berlin, Germany. Corresponding author: D. Grether, grether@vsp.tu-berlin.de.

Transportation Research Record: Journal of the Transportation Research Board, No. 2175, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 10–18.
DOI: 10.3141/2175-02

Overview

In MATSim, each traveler of the real system is modeled as an individual agent. The overall approach consists of three important parts:

- Each agent independently generates a so-called plan, which encodes its preferences during a certain time period, typically a day.
- All agents' plans are simultaneously executed in the simulation of the physical system, the so-called traffic flow simulation or mobility simulation.
- There is a mechanism that allows agents to learn. In the implementation, the system iterates between plan generation and traffic flow simulation. The system remembers several plans per agent and scores the performance of each plan. Agents normally choose the plan with the highest score, sometimes reevaluate plans with bad scores, and sometimes obtain new plans by modifying copies of existing plans.

A plan contains the itinerary of activities that the agent wants to perform during the day, plus the intervening trip legs the agent must take to travel between activities. An agent's plan details the order, type, location, duration, and other time constraints of each activity and the mode, route, and expected departure and travel times of each leg.

A plan can be modified by various modules. In the test scenario, the time adaptation module is used; the large-scale application additionally uses a router module. The time adaptation module changes the timing of an agent's plan. A very simple approach is used that applies a random "mutation" just to the duration attributes of the agent's activities (*9*). The router is a time-dependent best path algorithm using the link's generalized costs on the basis of the link travel times of the previous iteration (*10*). Mode choice will not be simulated by a module per se, but instead, by making sure that every agent has at least one "car" and at least one "public transit" plan (*11*).

One of the plans of every agent is marked as "selected." The traffic flow simulation executes all agents' selected plans simultaneously on the network and provides output describing what happened to each individual agent during the execution of its plan. The car traffic flow simulation is implemented as a queue simulation, in which each street (link) is represented as a first-in, first-out queue with two restrictions (*12, 13*). First, each agent has to remain for a certain time on the link, corresponding to the free speed travel time. Second, a link storage capacity is defined that limits the number of agents on the link; if it is filled up, no more agents can enter this link. The public transit simulation simply assumes that travel by public transit takes twice as long as travel by car on the fastest route in an empty network (*11*) and that the travel distance is 1.5 times the beeline distance. Public transit is assumed to run continuously and without capacity restrictions.

The modules base their decisions on the output of the traffic flow simulation (e.g., knowledge of congestion) by using feedback from the multiagent simulation structure (*14, 15*). This sets up an iteration cycle that runs the traffic flow simulation with specific plans for the agents and then uses the planning modules to update the plans; these changed plans are again fed into the traffic flow simulation until consistency between the modules is reached. The feedback cycle is controlled by the agent database, which also keeps track of multiple plans generated by each agent.

In every iteration, 10% of the agents generate new plans by taking an existing plan, making a copy of it, and then modifying the copy with the time adaptation or the router module. The other agents reuse one of their existing plans. The probability to change the selected plan is calculated according to

$$p_{\text{change}} = \min\left(1, \alpha \cdot e^{\beta \cdot (s_{\text{random}} - s_{\text{current}})/2}\right) \quad (1)$$

where

α = probability to change if both plans have the same score, set to 1%;
β = sensitivity parameter, set to 20 for tests and to 2 for large-scale Zurich simulations; and
$s_{\{\text{random,current}\}}$ = score of random and current plan (see later explanation).

In the steady state, this model is equivalent to the standard multinomial logit model:

$$p_j = \frac{e^{\beta \cdot s_j}}{\sum_i e^{\beta \cdot s_i}}$$

where p_j is the probability for plan j to be selected.

The repetition of the iteration cycle coupled with the agent database enables the agents to improve their plans over several iterations. Because the number of plans is limited for every agent by memory constraints, the plan with the worst performance is deleted when a new plan is added to a person who already has the maximum number of plans permitted. The iteration cycle continues until the system has reached a relaxed state. At this point, there is no quantitative measure of when the system is "relaxed"; the cycle is just allowed to continue until the outcome is stable.

Scoring Plans

To compare plans, it is necessary to assign a quantitative score to the performance of each plan. In this work, to be consistent with economic appraisal, a simple utility-based approach is used. The approach is related to the Vickrey bottleneck model (*16, 17*), but is modified to be consistent with the present approach on the basis of complete daily plans (*18, 19*). The elements of the approach are as follows:

- The total score of a plan is computed as the sum of individual contributions:

$$U_{\text{total}} = \sum_{i=1}^{n} U_{\text{perf},i} + \sum_{i=1}^{n} U_{\text{late},i} + \sum_{i=1}^{n} U_{\text{tr},i} \quad (2)$$

where

U_{total} = total utility for a given plan;
n = number of activities, which equals the number of trips (the first and the last activity—both "home"—are counted as one);
$U_{\text{perf},i}$ = (positive) utility earned for performing activity i;
$U_{\text{late},i}$ = (negative) utility earned for arriving late to activity i; and
$U_{\text{tr},i}$ = (negative) utility earned for traveling during trip i.

To work in plausible real-world units, utilities are measured in euros.

- A logarithmic form is used for the positive utility earned by performing an activity:

$$U_{\text{perf},i}\left(t_{\text{perf},i}\right) = \beta_{\text{perf}} \cdot t_{*,i} \cdot \ln\left(\frac{t_{\text{perf},i}}{t_{0,i}}\right) \tag{3}$$

where

t_{perf} = actual performed duration of the activity,
t_* = typical duration of an activity, and
β_{perf} = marginal utility of an activity at its typical duration.

β_{perf} is the same for all activities because in equilibrium all activities at their typical duration need to have the same marginal utility. $t_{0,i}$ is a scaling parameter that is related to the minimum duration and to the importance of an activity. As long as dropping activities from the plan is not allowed, $t_{0,i}$ has essentially no effect.

- The (dis)utility of being late is uniformly assumed as

$$U_{\text{late},i} = \beta_{\text{late}} \cdot t_{\text{late},i} \tag{4}$$

where β_{late} is the marginal utility (in euros/h) for being late, and $t_{\text{late},i}$ is the number of hours late to activity i. β_{late} is usually negative.

- The (dis)utility of traveling used in this paper is estimated from survey data, illustrated in Grether et al. (*20*). It is assumed to be dependent on the transport mode and the individual income as explained in the following section.

In principle, arriving early could also be punished. There is, however, no immediate need to punish early arrival because waiting times are already indirectly punished by forgoing the reward that could be accumulated by doing an activity instead (opportunity cost). In consequence, the effective (dis)utility of waiting is already $-\beta_{\text{perf}} t_{*,i}/t_{\text{perf},i} \approx -\beta_{\text{perf}}$. Similarly, that opportunity cost has to be added to the time spent traveling.

No opportunity cost needs to be added to late arrivals because the late arrival time is spent somewhere else. In consequence, the effective (dis)utility of arriving late remains at β_{late}.

INCOME-DEPENDENT UTILITY FUNCTION

There is some agreement that income effects play an important role in transport policy analysis [see, e.g., Franklin (*1*), Bates (*6*), Herriges and Kling (*21*), Kockelman (*22*), and Bates (*23*)]. The argument essentially is that monetary price changes affect different income groups differently. This is usually addressed by including income-dependent user preferences in the utility function.

FUNCTIONAL FORM

The functional form used for simulations is loosely based on Franklin (*1*) and is similar to Kickhöfer (*24*). A detailed derivation of this form and the estimation of the corresponding parameters are illustrated in Grether et al. (*20*). Hence, the utility functions of the two transport modes car and public transit (pt) are, according to Equation 2, given by

$$U_{\text{car},i,j} = +\frac{1.86}{\text{h}} t_{*,i} \cdot \ln\left(\frac{t_{\text{perf},i}}{t_{0,i}}\right) - \frac{1.52}{\text{h}} t_{\text{late},i} - 4.58 \frac{c_{i,\text{car}}}{y_j} - \frac{0.97}{\text{h}} t_{i,\text{car}}$$

$$U_{\text{pt},i,j} = +\frac{1.86}{\text{h}} t_{*,i} \cdot \ln\left(\frac{t_{\text{perf},i}}{t_{0,i}}\right) - \frac{1.52}{\text{h}} t_{\text{late},i} - 4.58 \frac{c_{i,\text{pt}}}{y_i} \tag{5}$$

The first summand refers to Equation 3 with $\beta_{\text{perf},i} = +1.86/\text{h}$; the second summand corresponds with Equation 4 and $\beta_{\text{late},i} = -1.52/\text{h}$. The third and fourth summands introduce mode and income dependency to the utility functions: y_i is the daily income of person j, and c_i is the monetary cost for the trip to activity i. The indices car and pt indicate the mode. Trip costs are calculated by using $c_{i,\text{car}} = 0.12$ CHF/km and $c_{i,\text{pt}} = 0.28$ CHF/km. Although there is a fourth summand for car ($\beta_{\text{tt,car}} = -0.97/\text{h}$), picking up the linear disutility of travel time t_i, there is no equivalent expression in the pt utility function. Travel time on pt is nonetheless punished by the opportunity costs of time by missing out on positive utility of an activity ($\beta_{\text{perf},i}$), which also implies additional negative utility for the car travel time.

The individual income is derived for each agent on the basis of several Lorenz curves. Methodical details can be found in Grether et al., also showing that a distribution similar to reality is generated (*20*). Utilities are computed in "utils"; a possible conversion into units of money or "hours of leisure time" needs to be done separately (see section on discussion of results) (*25*). Note that the late arrival parameter will be used only in the test, not in the real-world scenario.

TEST SCENARIO

The goal of this section is to verify the correctness and plausibility of the estimated choice model and the underlying implementation. Because probabilistic multiagent simulations and other software systems tend to be sensitive to new implementations, a simple setup is used to test the plausibility of traveler choice reactions as a result of a fictive policy change.

Network

The test network consists of a cycle of one-way links with (unrealistically) high capacities so as to minimize their influence on traffic patterns, essentially making it possible for most agents to drive with free speed. One link between home and work location has a reduced capacity of 1,000 vehicles per hour, building a bottleneck.

Initial Plans

The synthetic population consists of 2,000 agents. All agents start at their home activity, which they initially leave at 6:00 a.m. They initially drive to work with a car, stay there for 8 h and drive home afterward. The home-to-work trip has a length of 17.5 km, and the way back is 32.5-km long. Speed limit is at 50 km/h, so the free speed travel time from home to work by car is 21 min and 39 min are needed for the way back home. Thus the total free speed travel time driving by car is 60 min. Because the agents are forced to remain on that route, the scenario is similar to the well-known Vickrey bottleneck scenario (*16, 17*); also see below for more details.

In addition, each agent possesses an initially nonactive plan that uses the public transit mode for both trips. These trips take twice as

long as by car at free speed, that is, 42 min from home to work, and 78 min for the way back. The total public transit travel time is 120 min. In contrast to the car travel times, these transit travel times are not affected by congestion. Because public transit is assumed to run continuously and without capacity restrictions, a home departure at time t will always result in a work arrival at $t + 42$ min.

Work opens at 7:00 a.m. and closes at 6:00 p.m. To obtain the similarity to the Vickrey scenario, an additional behavioral parameter of $\beta_{late,i} = -1.52$/h is used, that is, deducting -1.52/h for arriving late. To be consistent with the Vickrey bottleneck scenario, any arrival time after 7:00 a.m. is directly considered as late. Estimation of income for the synthetic population is based on data from Kanton Zurich. The income distribution is retrieved from a Lorenz curve for 2006 in which the median income is 46,300 CHF (Swiss francs).

Behavioral Parameters

The behavioral parameters are set and can be interpreted as follows:

- Marginal utility of performing an activity at its typical duration: $\beta_{perf} = 1.86$/h,
- Marginal disutility of arriving late: $\beta_{late} = -1.52$/h,
- Marginal utility offset for traveling with a car: $\beta_{tr,car} = -0.97$/h,
- Marginal utility offset for traveling with public transit mode: $\beta_{pt} = 0$ (see below),
- Factor in logit process (Equation 1): $\beta = 20$, and
- Typical durations of $t_{*,w} = 8$ and $t_{*,h} = 12$ h for work and home, respectively, mean that work and home times have a tendency to arrange themselves with a ratio of 8:12 (i.e., 2:3). The time of the home activity is wrapped around, that is, a departure at 6 a.m. and a return at 5 p.m. result in a home activity duration of 13 h.

Work starts exactly at 7:00 a.m., meaning that (*a*) no utility can be accumulated from an arrival earlier than 7:00 a.m. and (*b*) any late arrival is immediately punished with $\beta_{late} = -1.52$/h.

Because of the argument made earlier concerning the opportunity cost of forgone activity time when arriving early, the effective marginal disutility of early arrival is $-\beta_{perf} t_{*,i}/t_{perf,i} \approx -\beta_{perf} = -1.86$/h, which is equal to the effective marginal disutility of traveling with public transit $\beta_{tt_{pt},eff}$. By the same argument, the effective marginal disutility of traveling by car is $\beta_{tt_{car},eff} = -\beta_{perf} t_{*,i}/t_{perf,i} - |\beta_{tt_{car}}| \approx -\beta_{perf} - |\beta_{tt_{car}}| \approx -2.83$/h. The return trip has no influence because there is no congestion.

Overall, the effective values of car travel time of the present study would correspond to the values $\beta_{early,eff}$, $\beta_{tt_{car},eff}$, $\beta_{late,eff} = (-1.86, -2.83, -1.52)$ of the Vickrey scenario (*16, 17*).

Simulation Runs

First, a preparatory run—referred to as base case—is performed for 4,000 iterations. During the first 2,000 iterations, 10% of the agents perform "time adaptation," that is, they make a copy of an existing plan and shift each element of its time structure by a random amount between zero and 7.5 min. The other 90% of agents switch between their existing plans according to Equation 1, which means that they potentially also switch the mode. During the second 2,000 iterations, time adaptation is switched off; in consequence, agents switch between existing plans only according to Equation 1. That is, their choice set now remains fixed to what they have found in the first 2,000 iterations, and they choose within this set according to a logit model.

After this, the policy measure is introduced. Every policy case is run for another 2,000 iterations, starting from the final iteration of the preparatory run. For the first 1,000 iterations of the policy case, the time adaption module is again switched on, with the same 10% replanning fraction. The final 1,000 iterations are once more with a fixed choice set. The following three different policy measures are simulated:

- Public transit price increase. The price of public transit, that is, is raised by 20% from 0.28 to 0.336 CHF/km.
- Public transit speed increase. The speed of public transit is increased, now taking only 1.8 (instead of 2.0) times as long as the free speed with car. This equals a speed increase of 10%.
- Combination. The two measures above are combined.

The policy design is based on the estimation about price and travel time elasticities (*26*). In this collection of different studies, travel time elasticities are identified to be double as high as price elasticities. Thus, for the combined policy measure, one would expect almost no shift in the modal split. In the following section, Iteration 4,000 of the base case is then compared with the final iteration of the combined measure. For reasons of clarity, results of the pure pt price increase and of the pure pt speed increase are not discussed in this paper. A detailed analysis can be found in Grether et al. (*20*).

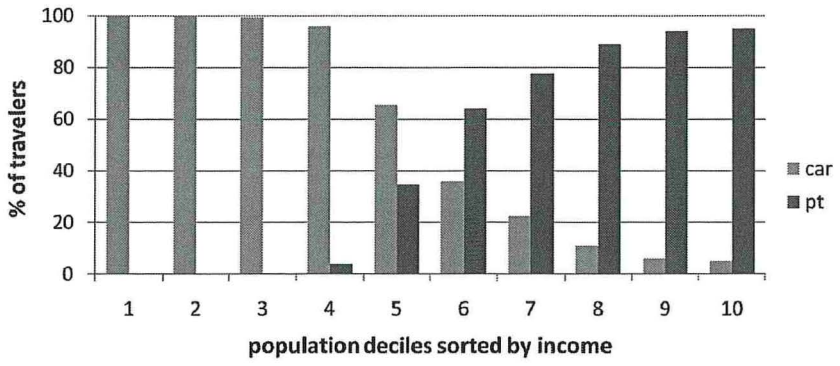
Results

According to Figure 1*a*, people with low incomes predominantly use the car, whereas people with high incomes predominantly use public transit. This means that car is the low value, and public transit, the high value mode. Even though that is a bit surprising, it is consistent with the higher costs of $c_{pt} = 0.28$ CHF/km assumed for public transit than for car ($c_{pt} = 0.12$ CHF/km), which were used for parameter estimation (*20, 24, 27*). The overall modal split of the base case is 54%:46% (car:pt).

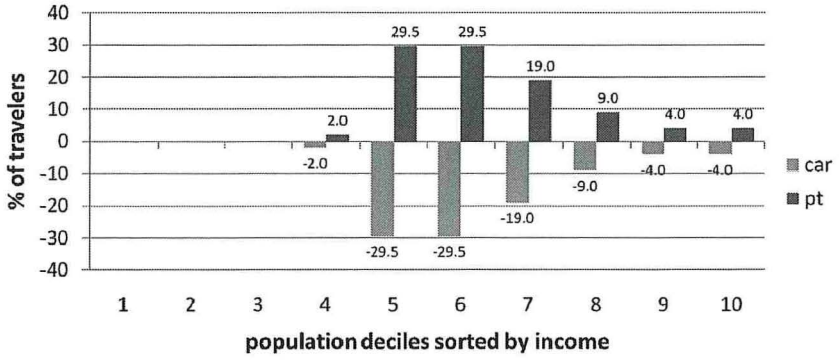
The combined measure leads to a modal split of 44%:56% (car:pt). Obviously, 10% of the agents change from car to pt. This happens through a shift of the income level that divides the two regimes: it moved toward lower incomes (Figure 1*b*).

Generally speaking, this means that people react more sensitively to the speed than to the price increase. These results are contradictory to the expectations based on Cervero that predict a roughly unchanged modal split (*26*).

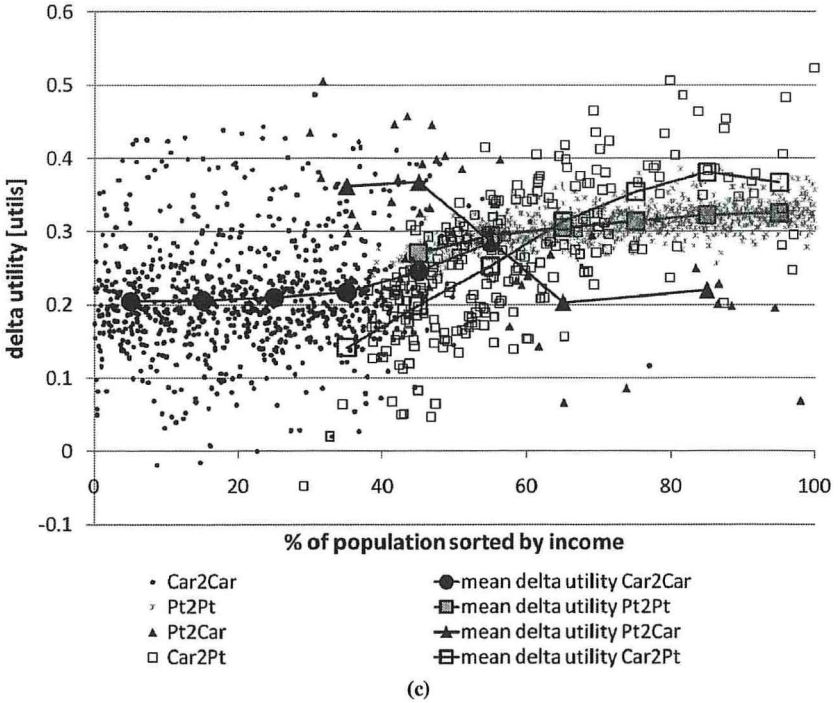
Figure 1*c* shows, agent by agent, the utility differences between the base case and the policy case as a scatter plot over deciles of the population. Every decile contains the same number of agents, sorted by their income. The four different geometric shapes in the plots correspond to four different user groups that can be identified as a result of the policy change. Dots and crosses represent agents that choose the same transport mode before and after the measure, a dot for the car mode, a cross for the pt mode. Triangles and blank quads represent agents that change their transport mode; a triangle means a change from pt to car and a blank quad means a switch from car to pt. For analysis purposes, mean values of utility change are computed for every group in the population deciles. A threshold of four was used for the plot, meaning that population deciles with fewer than four agents in the corresponding group were not taken into consideration.



(a)



(b)



(c)

FIGURE 1 Results of test scenario, modal split over deciles of the population sorted by income: (*a*) base case, absolute values (light gray bars depict car drivers, dark gray bars public transit users); (*b*) combined measure: changes in percentage points; and (*c*) combined measure: utility changes per person and mode (average utility changes per population decile sorted by income).

At first glance, one notices a "fan" of points at lower income levels and a more correlated structure stretching from middle to higher income levels. The fan can be traced back to the car users, who, because of stochastic congestion effects, face rather strong fluctuations of their utilities. In contrast, the correlated structure is caused by the pt users.

As Figure 1*c* shows, the measure leads to an overall utility gain across all income and user groups. First, public transit users predictably lose from the price increase, but simultaneously gain from the speed increase. The price increase component affects people with lower incomes more than people with higher incomes; the inverse is true for the speed increase. Second, a few mid-income people change from pt to car. Thus, for them the negative utility changes from the price increase overweight the travel time savings from the speed increase. Third, with increasing income the pt mode becomes more attractive because travel time savings become more important than the additional pt fee. Last, for car users, the pt price increase leads to an increased car share, thus more car congestion, thus utility losses for car users; however, the pt speed increase leads to a reduced car share, thus less car congestion, thus utility gains for car users. On average the car users still gain, which means that the global effect of the pt speed increase is dominating.

The second and the third effect are influenced strongly by stochastic effects in the plan selection process (see section on simulation structure): Mid-income agents that did, in the base case, randomly choose a pt plan instead of a better car plan, get on a higher utility level when changing to car after the policy only because of stochastic effects. At higher income levels, agents that in the policy case randomly choose a car plan instead of a better pt plan get on a lower utility level for the same reason. Vice versa, the same is true for the curve representing mean utility changes for agents that switch from car to public transport mode.

Overall, results demonstrate that the approach picks up the distributional effects of transport policy measures. Although both price and quality-of-service changes affect mode share, achieving this with price changes affects the lower-income groups more, whereas achieving this with quality-of-service changes affects the higher-income groups more. Thus, these plausibility tests can be considered successful. The approach is therefore applied to a real-world scenario of the Zurich metropolitan area in the next section.

SCENARIO: ZURICH, SWITZERLAND

The income-dependent utility function is now applied to a large-scale, real-world scenario. The area of Zurich, Switzerland, which has about 1 million inhabitants, is used. The following paragraphs give a simplified description of the scenario; a full description of the scenario and a calibration analysis can be found in Grether et al. (*20*) and Chen et al. (*28*).

Network and Population

The network is a Swiss regional planning network that includes the major European transit corridors. It consists of 24,180 nodes and 60,492 links.

The simulated demand consists of all travelers in Switzerland that are inside an imaginary 30-km boundary around Zurich at least once during their day. All agents have complete day plans with activities, such as home, work, education, shopping, and leisure, based on microcensus information (*29, 30*). The time window during which activities can be performed is limited to certain hours of the day: work and education can be performed from 7:00 a.m. to 6:00 p.m. and shopping from 8:00 a.m. to 8:00 p.m.; home and leisure have no restrictions. Each agent gets two plans based on the same activity pattern. The first plan uses only car as the transportation mode, and the second plan uses public transit only. Unlike the test scenario described above, there is no punishment for being late.

To speed up computations, a random 10% sample is taken from the synthetic population for simulation, consisting of 181,725 agents. In this large-scale application, the agents can, in addition to the time adaptation, perform route adaptation, which is essential for the car mode. Also mode adaptation is implicitly included (see section on simulation structure).

Simulation Runs

To maintain consistency with the test scenarios, the total number of iterations is reduced but the proportion of the different simulation steps is held constant. For the base case, that means

- For 1,000 iterations, 10% of the agents perform time adaptation and 10% adapt routes (see section on simulation structure). The rest of the agents switch between their existing plans, which implicitly include mode choice.
- During the second 1,000 iterations, time and route adaption are switched off; in consequence, agents switch only between existing options.

After this, the policy measure is introduced and is run for another 1,000 iterations, starting from the final iteration of the base case. Again, during the first 500 iterations, 10% of the agents perform time adaptation and another 10% of agents adapt routes. Agents adapting neither time nor route switch between existing plans and eventually switch between transport modes. For the final 500 iterations, only a fixed choice set is available.

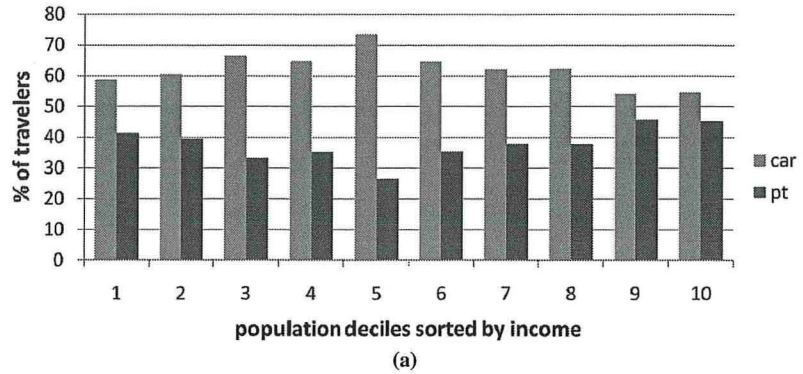
Different parameter combinations were tested, up to an overall 30% public transit speed increase and a 60% rise in public transit prices.

To evaluate the impact of the policy, Iteration 2,000 of the base case is now compared with the final iteration of the combined measure. Similar to the measure in the test scenario, public transit price is raised by 10% and speed is increased by 20%.

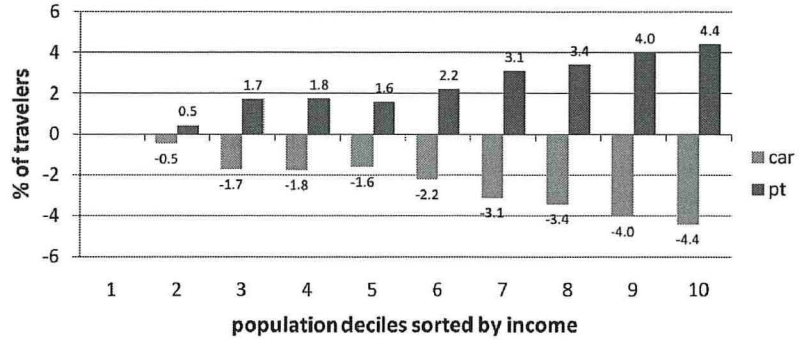
Results

The base case exhibits a modal split of 60.9%:39.1% (car:pt). Figure 2*a* depicts the modal split in the income deciles of the population. In contrast to the base case of the test scenario shown in Figure 1*a*, the distribution here is more homogeneous. Both modes are used across all deciles. The highest percentage of car users can be observed from the third to the fifth decile, whereas in the test scenario this is from the first to the third decile.

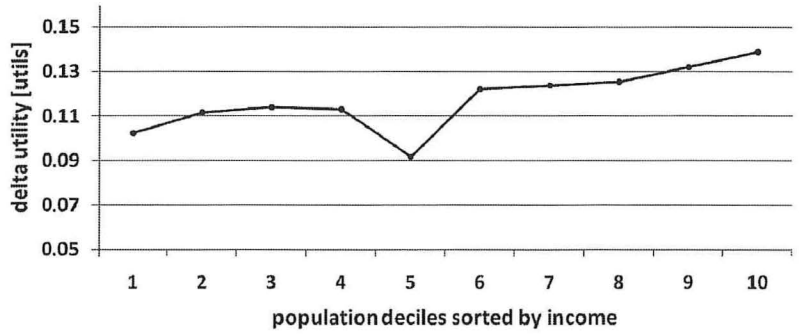
The combined measure for public transit results in a mode share of 58.5%:41.5% (car:pt). Because of the combined speed and price increase of pt, 2.4% of car travelers change from car to pt. Figure 2*b* compares mode share changes in the income deciles of the population to the base case. At a quick glance one can observe that with increasing income, more people switch from car to pt. More precisely, one



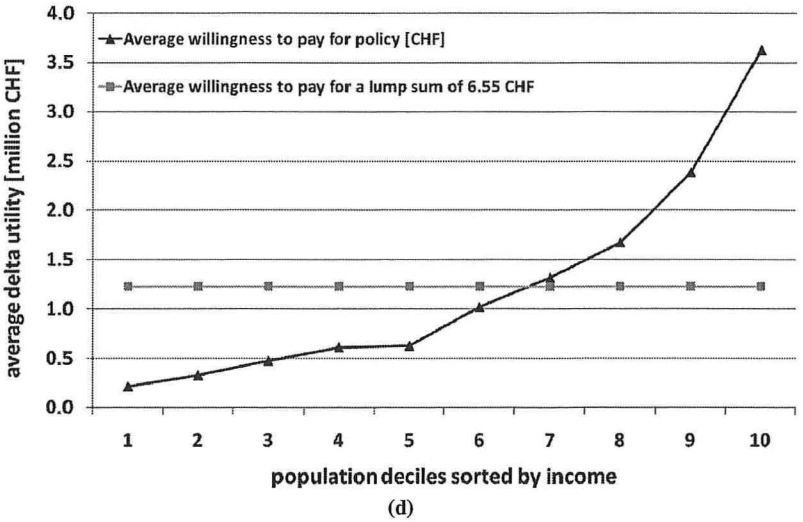
(a)



(b)



(c)



(d)

FIGURE 2 Results of Zurich scenario, modal split over income deciles: (*a*) base case (light gray bars depict car drivers, dark gray bars, public transit users); (*b*) combined measure versus base case, over deciles of population sorted by income; (*c*) average utility changes; and (*d*) daily willingness to pay for policy change.

can see a break in the increasing pt shares in the fifth decile, in which only 1.6% change mode, whereas in the fourth decile 1.8% change the transportation mode. Apart from this outlier the mode choice reflects the decreasing importance of travel costs compared with travel time savings when income increases as specified in Equation 5. This is more obvious than in the test scenario in which the strongest mode reaction to the measure takes place in the fifth and sixth deciles. Thereby the initial distribution of mode choice over income deciles should be taken into account. The test scenario's distribution is rather artificial, ranging from 100% car users in the first decile to 100% public transit users in the 10th decile. Exalting the incentive to switch from car to public transit for people already using this mode cannot show any effect in low-and high-income deciles. For this purpose the more uniform distribution of mode choice of the Zurich scenario is a more suitable starting point.

Increasing utility gains of agents with higher income can also be seen in Figure 2*c*, which depicts the average utility change of each population decile sorted by income. Each dot is located in the middle of the decile and represents the average utility change per decile. For representation purposes the dots are connected with lines. Obviously, one recognizes rising utility gains with increasing income. In regard to utils, the slope of the curve is slightly positive. The subsequent section shows that this increase has even stronger effects when utils are converted into money terms.

DISCUSSION OF RESULTS

For the combined policy case in the large-scale Zurich scenario, a basic analysis of welfare changes along deciles of the population is discussed. The overall effect is calculated by the mean utility gain in the deciles $\overline{\Delta U_d}$ (in regard to money) times the (always equal) number of persons in each decile n. According to Equation 5, conversion from utility units into CHF is based on individual income y_j and utility changes ΔU_j:

$$\overline{\Delta U_d} = \frac{1}{n} \sum_{j=1}^{n} \frac{\Delta U_j \cdot y_j}{4.58} \tag{6}$$

Summing this over all 10 deciles, the welfare effect of this policy is about 1.23 million CHF per day or almost 300 million CHF per year for the computed 10% sample of the Zurich metropolitan population (see section on Zurich, network and population). Thus, following standard economic appraisal methods, the policy should be introduced if this benefit outweighs economic costs.

Figure 2*d* shows in black the total daily monetarized gains over deciles of the population, sorted by income. The monetarized gains in every decile can be interpreted as the total willingness to pay for the measure. The gray curve tries to explain implementation problems due to low acceptance in the society. If, in a hypothetical case, the same daily welfare gains of 1.23 million CHF were distributed as a monetary lump-sum payment to every member of the population, every person would gain 6.55 CHF per day, or every decile 123,000 CHF. This highlights an important implementation problem of policy measures in democratically organized societies: almost 70% of the population would be better off with the lump-sum payment than with the implementation of the measure and are therefore likely to refuse the latter. Thus, if the simulation results are correct, financing this measure with tax revenues would be more appropriate, assuming a progressive income tax. Whereas financed by nondifferentiated user fees, this measure would have a regressive impact on the income distribution.

This example is meant to show some possibilities of economic policy evaluation that are feasible with multiagent microsimulations. Agents optimize their daily plans with respect to individual preferences, such as individual income or activity location. Still, there are three main issues that should be addressed in the future. First, for more reliable results, the survey should be designed in a way that all needed parameters can be estimated independently. Second, public transit is assumed to be 100% reliable, and no fluctuations due to geographic location or line cycles are considered. In principle, using multiagent transport simulations makes it possible to combine multiple demographic attributes of the population of interest, for example, by viewing the geospatial distribution of winners and losers of a measure (*5*). Neither the measure of this paper nor the public transit simulation features geospatial diversity. Thus analysis in the geographic dimension is strongly homogeneous, and a spatial pattern is not visible. In the case of a policy that is targeted on some geospatial impact, the multiagent approach should give interesting insights into geospatial distribution of gains and losses (*31*). Third, utility changes in the simulation are influenced by stochastic effects in the plan selection process, especially for people that switch mode. Nonetheless, it is shown that with this multiagent approach, welfare computations and equity analysis can be done on the desired level of (dis)aggregation.

CONCLUSION

Standard economic policy evaluation allows the realization of projects if the aggregated economic benefit overweights their costs. In democratically organized societies, the implementation of measures with regressive effects on the welfare distribution tends to be complicated because of low public acceptance.

The microscopic simulation approach presented in this paper is capable of helping in the design of better solutions in such situations. In particular, it is shown that income can be included in utility calculations for a better understanding of problems linked to acceptability. Then, in contrast to project evaluation applied in practice, choice modeling and economic evaluation are implemented in a consistent framework because the simulation values are used directly for evaluation. Furthermore, and going beyond Franklin (*1*), it is shown that the approach works in a large-scale, real-world example for which economic benefits are calculated.

ACKNOWLEDGMENTS

This work was funded in part by the Bundesministerium für Bildung und Forschung within the research project Adaptive Verkehrssteuerung (Advest) and in part by the German Research Foundation within the research project Detailed Evaluation of Transport Policies Using Microsimulation. The compute cluster is maintained by the Department of Mathematics, Technical University of Berlin.

REFERENCES

1. Franklin, J. P. *The Distributional Effects of Transportation Policies: The Case of a Bridge Toll for Seattle.* PhD dissertation. University of Washington, Seattle, 2006.

2. Pearce, D., and C. A. Nash. *Social Appraisal of Projects: A Text in Cost–Benefit Analysis.* Wiley & Sons, London, 1981.
3. Meister, K., M. Rieser, F. Ciari, A. Horni, M. Balmer, and K. W. Axhausen. Anwendung eines agentenbasierten Modells der Verkehrsnachfrage auf die Schweiz. *Proc., Heureka '08,* Stuttgart, Germany, March 2008.
4. Nagel, K., D. Grether, U. Beuck, M. Rieser, Y. Chen, and K. W. Axhausen. Multi-agent Transport Simulations and Economic Evaluation. Special issue on agent-based models for economic policy advice, (B. LeBaron and P. Winker, eds.) *Journal of Economics and Statistics (Jahrbücher für Nationalökonomie und Statistik),* Vol. 228, No. 2+3, 2008, pp. 173–194.
5. Grether, D., Y. Chen, M. Rieser, U. Beuck, and K. Nagel. Emergent Effects in Multi-Agent Simulations of Road Pricing. Presented at Annual Meeting of European Regional Science Association ERSA 2008, 2008.
6. Bates, J. Economic Evaluation and Transport Modelling: Theory and Practice. In *Moving Through Nets: The Physical and Social Dimensions of Travel* (K. W. Axhausen, ed.), Elsevier, Netherlands, 2006, pp. 279–351.
7. de Jong, G., A. Daly, E. Kroes, and T. van der Hoorn. Using the Logsum in Project Appraisal. *Proc., Meeting of International Association for Travel Behavior Research,* Kyoto, Japan, 2006. www.iatbr.org.
8. Raney, B., and K. Nagel. An Improved Framework for Large-Scale Multi-Agent Simulations of Travel Behavior. In *Towards Better Performing European Transportation Systems* (P. Rietveld, B. Jourquin, and K. Westin, eds.), Routledge, London, 2006, pp. 305–347.
9. Balmer, M., B. Raney, and K. Nagel. Adjustment of Activity Timing and Duration in an Agent-Based Traffic Flow Simulation. In *Progress in Activity-Based Analysis* (H. J. P. Timmermans, ed.), Elsevier, Oxford, United Kingdom, 2005, pp. 91–114.
10. Lefebvre, N., and M. Balmer. Fast Shortest Path Computation in Time-Dependent Traffic Networks. *Proc., Swiss Transport Research Conference.* www.strc.ch.
11. Rieser, M., D. Grether, and K. Nagel. Adding Mode Choice to a Multi-Agent Transport Simulation. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2132,* Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 50–58.
12. Gawron, C. *Simulation-Based Traffic Assignment.* PhD dissertation. University of Cologne, Germany, 1998.
13. Cetin, N., A. Burri, and K. Nagel. A Large-Scale Agent-Based Traffic Microsimulation Based on Queue Model. *Proc., Swiss Transport Research Conference.* www.strc.ch.
14. Kaufman, D. E., K. E. Wunderlich, and R. L. Smith. *An Iterative Routing/Assignment Method for Anticipatory Real-Time Route Guidance.* Technical Report IVHS. Technical Report 91-02. Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, May 1991.
15. Bottom, J. *Consistent Anticipatory Route Guidance.* PhD dissertation. Massachusetts Institute of Technology, Cambridge, Mass., 2000.
16. Vickrey, W. S. Congestion Theory and Transport Investment. *The American Economic Review,* Vol. 59, No. 2, 1969, pp. 251–260.
17. Arnott, R., A. de Palma, and R. Lindsey. Economics of a Bottleneck. *Journal of Urban Economics,* Vol. 27, No. 1, 1990, pp. 111–130.
18. Charypar, D., and K. Nagel. Generating Complete All-Day Activity Plans with Genetic Algorithms. *Transportation,* Vol. 32, No. 4, 2005, pp. 369–397.
19. Raney, B., and K. Nagel. An Improved Framework for Large-Scale Multi-Agent Simulations of Travel Behaviour. In *Towards Better Performing European Transportation Systems* (P. Rietveld, B. Jourquin, and K. Westin, eds.), Routledge, London, 2006, pp. 305–347.
20. Grether, D., B. Kickhöfer, and K. Nagel. Policy Evaluation in Multi-Agent Transport Simulations Considering Income-Dependent User Preferences. *Proc., 12th Conference of International Association for Travel Behaviour Research,* Jaipur, India, Dec. 13–18, 2009.
21. Herriges, J. A., and C. L. Kling. Nonlinear Income Effects in Random Utility Models. *The Review of Economics and Statistics,* Vol. 81, No. 1, 1999, pp. 62–72.
22. Kockelman, K. M. A Model for Time- and Budget-Constrained Activity Demand Analysis. *Transportation Research Part B,* Vol. 35, No. 3, 2001, pp. 255–269.
23. Bates, J. Measuring Travel Time Values with a Discrete Choice Model: A Note. *Economic Journal,* Vol. 97, No. 386, 1987, pp. 493–98.
24. Kickhöfer. B. *Die Methodik der ökonomischen Bewertung von Verkehrsmaßnahmen in Multiagentensimulationen.* MS thesis. Technische Universität Berlin, 2009.
25. Jara-Díaz, S., M. Munizaga, P. Greeven, R. Guerra, and K. W. Axhausen. Estimating the Value of Leisure from a Time Allocation Model. *Transportation Research B,* Vol. 42, No. 10, 2008, pp. 946–957.
26. Cervero, R. Transit Pricing Research. *Transportation,* Vol. 17, 1990, pp. 117–139.
27. Vrtic, M., N. Schüssler, A. Erath, M. Bürgle, K. W. Axhausen, E. Frejinger, M. Bierlaire, R. Rudel, S. Scagnolari, and R. Maggi. Einbezug der Reisekosten bei der Modellierung des Mobilitätsverhaltens. Schriftenreihe 1191, Bundesamt für Strassen, UVEK, Bern, Switzerland. 2008.
28. Chen, Y., M. Rieser, D. Grether, and K. Nagel. Improving a Large-Scale Agent-Based Simulation Scenario. VSP Working Paper 08-15. VSP, Technische Universität Berlin, 2008. www.vsp.tu-berlin.de/publications.
29. Eidgenössische Volkszählung. Swiss Federal Statistical Office, Neuchatel, Switzerland, 2000.
30. Ergebnisse des Mikrozensus 2005 zum Verkehr. Swiss Federal Statistical Office, Neuchatel, Switzerland, 2006.
31. Rieser M., and K. Nagel. Combined Agent-Based Simulation of Private Car Traffic and Transit. *Proc., 12th Conference of International Association for Travel Behaviour Research,* Jaipur, India, Dec. 13–18, 2009.

The Transportation Demand Forecasting Committee peer-reviewed this paper.

Travel Time Forecasting and Dynamic Origin–Destination Estimation for Freeways Based on Bluetooth Traffic Monitoring

Jaume Barceló, Lidin Montero, Laura Marqués, and Carlos Carmona

Traditional technologies, such as inductive loop detectors, do not usually produce measurements of the quality required by real-time applications. Therefore, one wonders what could be expected from newer information and communication technologies, such as automatic vehicle location, license plate recognition, and detection of mobile devices. The main objectives of this paper are to explore the quality of the data produced by Bluetooth detection of mobile devices that equip vehicles for travel time forecasting and its use in estimating time-dependent origin–destination matrices. Ad hoc procedures based on Kalman filtering have been designed and implemented successfully, and the numerical results of the computational experiments are presented and discussed.

The objective of this paper is to explore the design and implementation of methods that support the short-term forecasting of expected travel times and to estimate the time-dependent origin–destination (O-D) matrices with new detection technologies. This is the case for new sensors that detect vehicles equipped with Bluetooth mobile devices, that is, hands-free phones, Tom-Tom, Parrot, and similar devices, whose penetration is becoming ever more pervasive. From a research point of view, this means starting to explore the potential of a new technology that improves traffic models and at the same time provides practitioners with sound applications that are easy to implement.

Concerning the information supplied by an advanced traveler information system (ATIS) to motorists entering a freeway, there is a wide consensus in considering forecast travel time as one of the most useful inputs from a driver's perspective. Forecast travel time is the expected travel time experienced when traversing a freeway segment, rather than the instantaneous travel time, which is the travel time of a vehicle traversing a freeway segment at time t if all traffic conditions remain constant until the vehicle exits the freeway. That usually underestimates or overestimates travel time, depending on traffic conditions. The same applies to reconstructed travel time (the travel time realized at time t when a vehicle leaves a freeway segment), which represents a past travel time. See, for instance, Travis et al. (*1*).

Department of Statistics and Operations Research and Center for Innovation in Transport, Technical University of Catalonia, Campus Nord, Building C5, Office 215, 08034 Barcelona, Spain. Corresponding author: J. Barceló, jaume.barcelo@upc.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2175, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 19–27.
DOI: 10.3141/2175-03

The dynamic estimates of time dependencies in O-D matrices are a major input to dynamic traffic models used in an advanced traffic management system (ATMS) for estimating the current traffic state as well as for forecasting its short-term evolution. Travel time forecasting and dynamic O-D estimation are thus two of the key components of ATIS and ATMS, and the quality of the results that they can provide depends on the quality of the models as well as on the accuracy and reliability of the traffic measurements of traffic variables supplied by the detection technology.

Traditional technologies, such as inductive loop detectors, do not usually produce measurements of the quality required by real-time applications; therefore one wonders what could be expected from the new information and communication technologies, such as automatic vehicle location, license plate recognition, and detection of mobile devices. Consequently, the objectives of this paper are to explore methods for estimating time-dependent O-D matrices and short-term travel time forecasting on the basis of data produced by the detection of Bluetooth mobile devices on board vehicles.

CAPTURING TRAFFIC DATA WITH BLUETOOTH SENSORS

The sensor integrates a mix of technologies that enable it to audit the Bluetooth and Wi-Fi spectra of devices within its coverage radius. It captures the public parts of the Bluetooth or Wi-Fi signals. Bluetooth is the global standard protocol (IEEE 802.15.1) for exchanging information wirelessly between mobile devices by using a 2.4-GHz short-range radio frequency bandwidth. The captured code consists of a combination of six alphanumeric pairs (hexadecimal). The first three pairs are allocated to the manufacturer (Nokia, Panasonic, Sony, etc.) and the type of device (i.e., phone, hands free, Tom-Tom, Parrot, etc.), as allocated by IEEE; the last three define the MAC address, a unique 48-bit address assigned to each wireless device by the service provider company. The uniqueness of the MAC address makes it possible to use a matching algorithm to log the device when it becomes visible to the sensor. The logged device is time-stamped, and when it is logged again by another sensor at a different location, the difference in time stamps can be used to estimate the travel time between the locations. A vehicle equipped with a Bluetooth device traveling along the freeway is logged and time-stamped at time t_1 by the sensor at Location 1. After traveling a certain distance it is logged and time-stamped again at time t_2 by the sensor at Location 2 downstream. The difference in time stamps $\tau = t_2 - t_1$ measures the travel time of the vehicle equipped with that mobile device. Obviously the

speed is also measured, assuming that the distance between the locations is known. Data captured by each sensor are sent for processing to a central server by General Packet Radio Service.

Raw measured data cannot be used without preprocessing, which is aimed at filtering out outliers that could bias the sample, for example, a vehicle that stops at a gas station between sensor locations. To remove these data from the sample, a filtering process consisting of an adaptive mechanism has been defined; it assumes a lower-bound threshold for the free-flow speed v_f in that section estimated by previous traffic studies, which defines an upper bound τ_f to the travel time between sensors at 1 and 2 in these conditions. Travel times larger than that threshold are removed as abnormal data. The system monitors every minute of the aggregated average speed of the detected vehicles and updates the thresholds accordingly by increasing or decreasing the value step-by-step to match the measured values when they get close to the threshold. If the system is unable to generate any match in more than 2 min, the range is open to a maximum time value (5 km/h).

Because this sensor system can monitor the path of a vehicle, questions about the privacy of drivers could arise. However, working with the MAC address of Bluetooth devices ensures privacy because the MAC address is not associated with any other personal data; the audited data cannot be related to particular individuals. Besides, so as to reinforce the security of data, an asymmetric encryption algorithm is applied before data leave the sensor and get to the database, making it impossible to recover the original data (*2*).

TRAVEL TIME MEASUREMENTS AND FORECASTS

A pilot project has been conducted north of Barcelona, Spain, on a 40-km-long section of the AP-7 Motorway, between Barcelona and the French border. Figure 1 maps the area of the pilot project and highlights the motorway length and, by using colored circles, the location of the sensors involved, which are positioned on mileposts at Kilometers 87.2, 91.3, 106.4, 119.2, 125.4, and 130.5 of the AP-7 Motorway.

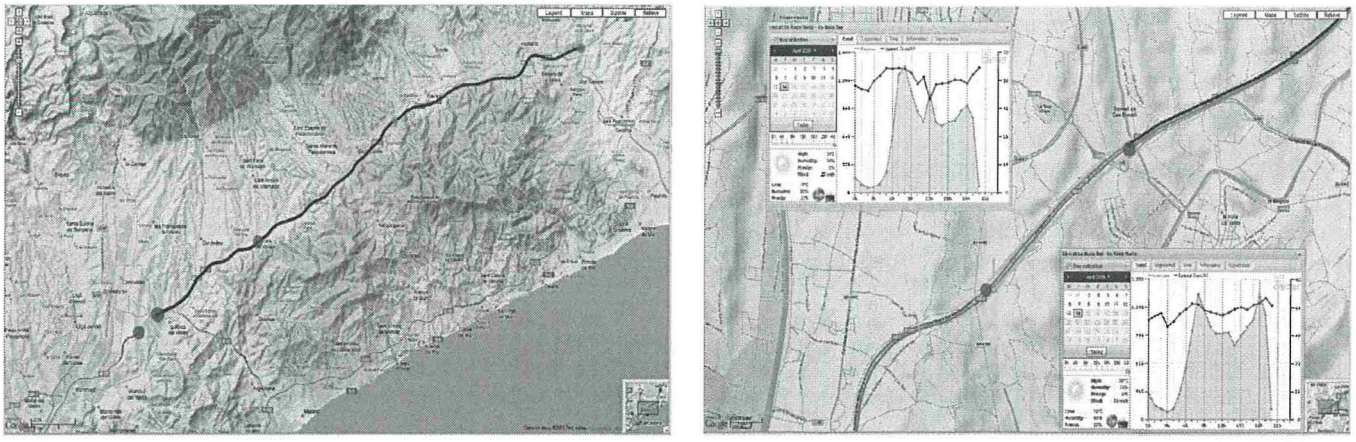
Figure 1 also depicts two examples of the measurements provided by sensors at the borders of a motorway segment. The upper graphic corresponds to southward flow, and the lower graphic corresponds to northward flow. The black line displays the time evolution of the speed between both locations throughout the day, and the blue area displays the quantity of detected devices. Tables 1 and 2 show an example of the raw data collected by the Bluetooth sensor and the filtered data used for forecasting. The ID column of Table 1 identifies the temporal identity assigned by the encryption algorithm. Time 1 and Time 2 identify the time stamps. The final two columns correspond to the calculated speed and travel time.

TABLE 1 Examples of Raw Measured Travel Times ($t_2 - t_1$) and Speeds

ID	Time 1	Time 2	km/h	t_2–t_1 (s)
10483	11-06-2009 19:07	11-06-2009 19:24	149.24	989
11925	11-06-2009 18:29	11-06-2009 18:47	133.33	1,107
12660	11-06-2009 18:48	11-06-2009 19:06	134.92	1,094
18419	11-06-2009 17:18	11-06-2009 17:40	113.89	1,296
18613	11-06-2009 19:35	11-06-2009 19:53	136.16	1,084

Data were collected during a 2-month period, May and June 2009, and were used to create a historical database of past measurements and traffic patterns which, together with real-time detection, provided the input for the forecasting algorithm.

Estimation and short-term prediction of travel times are a key component of ATIS. Consequently, they have attracted the interest of researchers in recent years. There have been a significant number of contributions dealing with various methods to achieve acceptable forecasting when measurements come from inductive loop detectors; these methods are based mostly on applications of traffic flow theory. Other researchers have drawn their attention to cases in which data are supplied by other technologies such as probe vehicles (*3*), or when cell phones or electronic toll identifications are the data sources (*4, 5*). In all these cases Kalman filtering has been proposed as the forecasting technique (*6*). It is an iterative process for modeling the evolution over time of a dynamic stochastic system that makes a prediction of the expected state of the system based on an estimate of the current state and the available measurements. If system S is in state E_{k-1} at



(a) (b)

FIGURE 1 Pilot project in Barcelona: (*a*) site of AP-7 Motorway in Barcelona and (*b*) two examples of Bluetooth detection: speeds and quantity of detected devices.

TABLE 2 Example of Filtered Data, Aggregated by Minute

Time	Total	Phones	Cars	Travel Time (s)	Speed (km/h)
11-06-2009 17:00	4	0	4	1,342	112.23
11-06-2009 17:01	8	4	4	1,400	109.93
11-06-2009 17:02	2	1	1	1,282	115.19
11-06-2009 17:03	7	5	2	1,508	100.84
11-06-2009 17:04	4	3	1	1,403	107.46

time $k-1$, defined by the values of the state variables $x(k-1)\in\Re$ at that time, then the values of the state variables change over time according to a dynamic process modeled by a transition equation. This transition equation defines the transition from state E_{k-1} to state E_k, and it is usually formulated in relation to the stochastic linear equation in differences:

$$x(k)=A(k-1)x(k-1)+w_k \tag{1}$$

where $A(k-1)$ is the transition function at time $k-1$ that captures the dynamics of the process and w_k is an error term representing the process error whose probability distribution is normal with zero mean and covariance Q, $[P(w) \sim N(0, Q)]$. Kalman filtering predicts the state $x_p(k)$ at time k from the transition equation at the previous time interval $k-1$ and the estimated state $x_e(k-1)$ at $k-1$:

$$x_p(k)=A(k-1)x_e(k-1)+w_k \tag{2}$$

To estimate the system's state Kalman filtering assumes that a measurement $z(k)$ is available and is related to the state by the linear relationship $z(k)=Hx(k)+v_k$, where H is the measurement function; the measurement equation is also affected by a measurement error v_k with a normal probability distribution of zero mean and covariance R, $[P(v) \sim N(0, R)]$. Then the a posteriori estimate of the state $x(k)$ in regard to the current measurement and the predicted measurement is formulated as

$$x_e(k)=x_p(k)+K(k)\left[z(k)-x_p(k)\right] \tag{3}$$

where factor $K(k)$ is called the "Kalman gain" and is the value that minimizes the covariance of the error of the a posteriori estimation in relation to the covariance $P_p(k)$, of the a priori error $\epsilon(k)=x(k)-x_p(k)$. The Kalman gain is given by

$$K(k)=P_p(k)H^T\left[HP_p(k)H^T+R(k)\right] \tag{4}$$

To complete the process all that is needed is to estimate the covariance $P_p(k)$ in relation to the covariance error $P_e(k-1)$ and the covariance of the process noise Q, which is done by

$$P_p(k)=A(k-1)P_e(k-1)A(k-1)+Q(k) \tag{5}$$

where the update of the covariance error $P_e(k)$ is given by

$$P_e(k)=\left[1-K(k)\right]P_p(k) \tag{6}$$

The Kalman filtering algorithm iterative process between prediction and updating is based on measurements whose main iterative steps are as follows:

1. Calculate the Kalman gain $K(k)$.
2. Update the measurements $z(k)$.
3. Calculate the a priori estimate of $x_p(k)$.
4. Update the covariance of the a posteriori error $P_e(k)$.

This can be formalized in regard to the following generic algorithm:

Step 0. Initialization.

Set $k=0$, $A(0)=1$, and $P(0)=\mathrm{Var}\left[\hat{z}(0)\right]$

$N=$ number of time intervals

Step 1. State prediction and measurement error covariance estimate.

$$x_p(k)=A(k-1)x_e(k-1)+w_k$$
$$P_p(k)=A(k-1)P_e(k-1)A(k-1)+Q(k)$$

Step 2. Kalman gain calculation.

$$K(k)=P_p(k)H^T\left[HP_p(k)H^T+R(k)\right]$$

Step 3. State estimate.

$$x_e(k)=x_p(k)+K(k)\left[z(k)-x_p(k)\right]$$

Step 4. Measurement error covariance update.

$$P_e(k)=\left[1-K(k)\right]P_p(k)$$

Step 5. If $k=N$, stop.

Otherwise, set $k=k+1$, and repeat from Step 1.

If the state variable $x(k)$ is the average travel time between two sensors, then the application of this algorithm to travel time forecasting based on Bluetooth measurements can be simplified, assuming that the measurement function H is equal to the identity matrix $z(k)=x(k)+v_k$, which is the measured travel time at time period k. The transition function $A(k)$ is given by

$$A(k)=\frac{\hat{z}(k)}{\hat{z}(k-1)} \tag{7}$$

where

$\hat{z}(k)$ = average historical travel times in the database for time period k for the traffic patterns corresponding to that time period,
Q = zero, and
$R(k)$ = estimated from the travel time variances of the corresponding traffic patterns in the database.

ESTIMATION OF TIME-DEPENDENT O-D MATRICES

Data Collection for Estimating Time-Dependent O-D

The possibility of tracking vehicles equipped with Bluetooth mobile technology raises the question of whether this information can be used for estimating the time-dependent O-D matrix whose entries $T_{ij}(k)$ represent the number of vehicles accessing the freeway at time interval k, using entry ramp i with exit ramp j as their destination.

A simulation experiment has been conducted before deploying the technology for a pilot project. The selected site is an 11.551-km-long section of the Ronda de Dalt, an urban freeway in Barcelona between the Trinitat and the Diagonal exchange nodes. The site has 11 entry ramps and 12 exit ramps (including main section flows) in the studied section, whose direction is toward Llobregat (to the south of the city). Figure 2 depicts a part of the site with the suggested sensor layout. D_i denotes the location of the ith sensor at the main section; E_j denotes the sensor located at the jth entry ramp, and S_n is the sensor located at the nth exit ramp. Vehicles are generated randomly in the simulation model according to a selected probability distribution, that is, an exponential shifted time headway, whose mean has been adjusted to generate the expected mean $T_{ij}(k)$ of vehicles for each O-D pair (i, j) at each time interval k. Once a vehicle is generated, it is randomly identified as an equipped vehicle depending on the proportion of penetration of the technology, 30% in this case, according to the available information on the penetration of the technology in the metropolitan region of Barcelona. The simulation emulates the logging and time-stamping of this random sample of equipped vehicles. Sensors are located at each entry and exit ramp and in the main stream immediately after each ramp.

Bluetooth and Wi-Fi data are collected every second and are matched when the same emulated MAC address is detected by sensors at entry ramps, exit ramps, and main sections; this provides the corresponding counts for each time interval. As a result, travel times between detectors are obtained in a way similar to that shown in Figure 1. Bluetooth and Wi-Fi sensors provide traffic counts and travel times between pairs of sensors for any time interval up to 0.1 s for equipped vehicles. The measured travel times at each time interval are as follows:

- Travel times from each on-ramp entering the corridor to every off-ramp exiting the corridor and
- Travel times from each on-ramp entering the corridor to every main section where a sensor has been installed.

Taking into account that Bluetooth sensors are tagging and time-stamping vehicles entering the motorway via entry ramp i at time interval k and tagging them again when they are leaving the motorway via exit ramp j, then Bluetooth detection is generating a sample $T_{ij}(k)$ of the number of vehicles entering the motorway at i during time interval k and later leaving at j. Therefore, it is natural to consider expanding the sample to the whole population to estimate the time-dependent O-D matrix $T_{ij}(k)$. This is a question that deserves further research. Comparing the number of detected Bluetooth-equipped vehicles with the number of vehicles counted by well-calibrated inductive loops located at the same position and taking the inductive loop sample as a reference, it was found that the variability of the Bluetooth sample yielded unacceptable expansion errors. Despite the fact that there was a high correlation between the two samples and that the variability of the Bluetooth sample matched quite well with the reference sample, these expansion errors invalidated any simple expansion procedure.

Consequently, it is still risky to make a straightforward estimation of O-D matrices based only on Bluetooth counting of vehicles, but the accuracy in measuring speeds and travel times opens the door to more efficient possibilities of using Kalman filtering for O-D estimates, simplifying the equations, and replacing state variables with measurements, as described in the next section.

Kalman Filter Approach for Estimating Time-Dependent O-D Matrices

The estimation of O-D matrices from traffic counts has received a great deal of attention in the past decades. The extension to dynamic O-D estimation in a dynamic system environment from time-series traffic counts has been frequently proposed [see Van Der Zijpp and

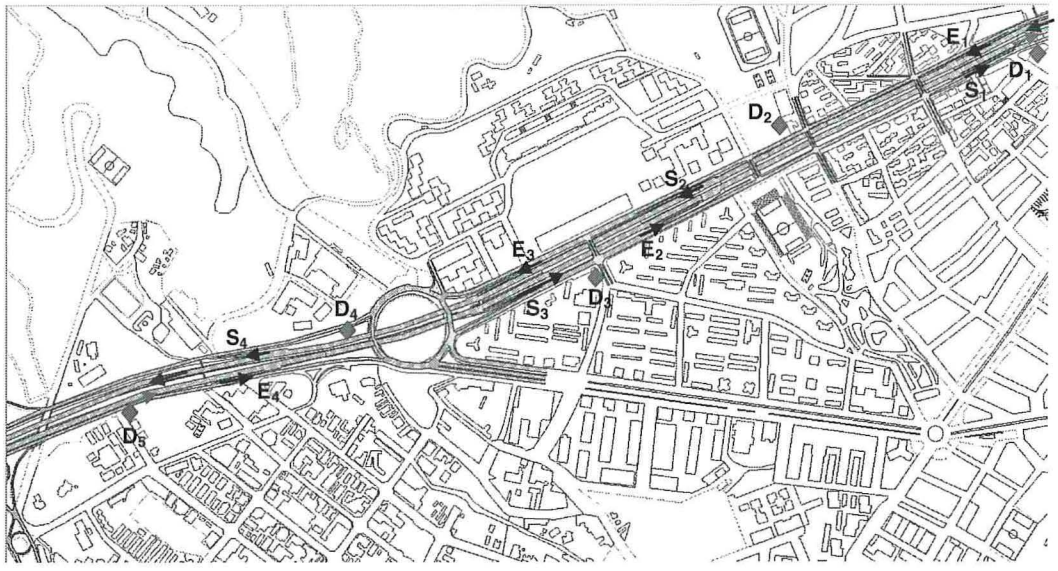


FIGURE 2 Segment of site for O-D estimation showing part of the detection layout.

Hamerslag (*7*) and Chang and Wu (*8*) and more recently Chu et al. (*9*) and Work et al. (*10*)]. A review of the studies up to 1991 is available in Bell (*11*).

The system equations for O-D estimation from static counts are underdetermined because there are far more O-D pairs than the existing number of equations. However, because dynamic methods employ time-series traffic counts, then the number of equations is larger than the O-D number, and a unique O-D matrix can thus be obtained. In the static and the dynamic methods, the relationships between O-D matrices and traffic counts must usually be defined in regard to an assignment matrix. Static methods, usually specialized for urban networks, establish the relationships between O-D pairs and link flows through static traffic assignment models embedded into entropy or bilevel mathematical programming models, depending on the approach (*12*). The availability of multiple alternative paths for each O-D pair is the crucial difference between linear networks (i.e., freeways) and more complex network topologies (i.e., urban networks). Thus, route choice becomes a key component in this case, and therefore the estimates are formulated in regard to the proportions of O-D flows using each of the available paths. Approaches are usually then based on an underlying dynamic traffic assignment and are the object of research [for details, see Ben-Akiva et al. (*13*) and Mahmassani and Zhou (*14*)].

This research has been oriented toward the dynamic O-D estimation in linear congested corridors (without route choice strategies, because there is only a unique path), taking into account travel times between O-D pairs affected by congestion. If no congestion exists but there is a constant delay for each O-D pair, the problem can be solved by any of the methods proposed by Bell (*11*), Van Der Zijpp and Hamerslag (*7*), or Chan and Wu (*8*).

Bell formulates a space–state model and applies the Kalman filter, taking into consideration a fixed and nonnegligible O-D travel time distribution for each O-D pair whereas no counts on the main section are considered (*11*). Stability in traffic conditions is needed during the estimation process, and congestion that arises cannot be captured by the formulation.

Van Der Zijpp and Hamerslag proposed a space–state model assuming for each O-D pair a fixed and nonnegligible O-D travel time distribution (*7*). The state variables are time-varying O-D proportions (between an entry and all possible destination ramps). The observation variables are main section counts for each interval. No exit ramp counts are available, and the relationship between the state variables and the observations includes a linear transformation that explicitly accounts for the number of departures from each entry during time interval k as well as a constant indicator matrix that details O-D pairs intercepted by each section detector. Suggestions for dealing with structural constraints on state variables are proposed. The Kalman filter process is interpreted as a Bayesian estimator, of which the initialization and noise properties are widely discussed. Tests with simulated data were conducted by comparing several methods, and the Kalman filter was reported to have performed better than the others. Fixed O-D travel time delays are not clearly integrated into the space–state model, although that is considered in some respect by the authors.

Chang and Wu proposed a space–state model that considers for each O-D pair a nonfixed O-D travel time estimated from time-varying traffic measures (*8*). Traffic flow models are implicitly included in the state variables. The state variables are time-varying O-D proportions and fractions of O-D trips that arrive at each off-ramp m intervals after their entrance at interval k. The observation variables are main section, and off-ramp counts for each interval and the relationship between the state variables and the observations are complex and nonlinear. An extended Kalman filter approach is proposed, and two algorithmic variants are implemented, one of them well suited for online applications.

Work et al. propose the use of an ensemble Kalman filtering approach as a data assimilation algorithm for a new highway velocity model proposal based on traffic data from Global Positioning System-enabled mobile devices (*10*).

A space–state formulation for dynamic O-D matrix estimation in corridors is proposed, considering congestion in such a way that combines elements of the Chang and Wu (*8*) and Van Der Zijpp and Hamerslag proposals (*7*). A linear Kalman-based filter approach is implemented for recursive state variable estimation. Tracking of the vehicles is undertaken by processing Bluetooth and Wi-Fi signals whose sensors are located as described above. Traffic counts for every sensor and O-D travel time from each entry ramp to the other sensors (main section and ramps) are available for any selected time interval length greater than 1 s. Travel time delays between O-D pairs or between each entry and sensor locations are provided directly by the detection layout and are no longer state variables but measurements that simplify the approach and make it more reliable.

A basic hypothesis that requires a statistical contrast for test site applications, and the authors assume holds true in what follows, is that equipped and nonequipped vehicles are assumed to follow common O-D patterns. Time interval length between 1 and 3 min is suggested to be able to detect arising congestion. Consider a corridor section containing ramps and sensors numbered as in Figure 2. The notation is defined below:

$q_i(k)$ = number of equipped vehicles entering the freeway from on-ramp i during interval k $i = 1, \ldots, I$;
$s_j(k)$ = number of equipped vehicles leaving freeway by off-ramp j during interval k $j = 1, \ldots, J$;
$y_p(k)$ = number of equipped vehicles crossing main section sensor p and $p = 1, \ldots, P$;
$G_{ij}(k)$ = number of vehicles entering the freeway at on-ramp i during interval k with destination to off-ramp j;
$g_{ij}(k)$ = number of equipped vehicles entering freeway from ramp i during interval k that are headed toward off-ramp j;
IJ = $I \times J$ = number of feasible O-D pairs;
$t_{ij}(k)$ = average measured travel time for equipped vehicles entering from entry i and leaving by off-ramp j during interval k;
$t_{ip}(k)$ = average measured travel time for equipped vehicles entering from entry i and crossing sensor p during interval k;
$b_{ij}(k)$ = $g_{ij}(k)/q_i(k)$, the proportion of equipped vehicles entering the freeway from ramp i during interval k that are destined to off-ramp j;
$U^h_{ijq}(k)$ = 1 if average measured time-varying travel time during interval k for traversing freeway section from entry i to sensor q takes h time intervals, where $h = 1, \ldots, M$, $q = 1, \ldots, Q$, and $Q = J + P$ (the total number of main section and off-ramp sensors), and M is the maximum number of time intervals required by vehicles to traverse entire freeway section considering a high-congestion scenario;
= 0 otherwise;
$e(k)$ = e = fixed column vector of dimension I containing ones; and
$z(k)$ = observation variables during interval k; i.e., a column vector of dimension $I + J + P$, whose structure is $z(k)^T = [s(k)\ y(k)\ e(k)]^T$.

The state variables are time-varying O-D proportions for equipped vehicles entering the freeway from ramp i during interval k with

destination off-ramp j. The observation variables are main section and off-ramp counts for each interval k. The relationship between the state variables and the observations involves a time-varying linear transformation that considers the following:

- Number of equipped vehicles entering from each entry during time interval k, $q_i(k)$ and
- M time-varying indicator matrices $[U_{ijq}^h(k)]$ detailing O-D pairs intercepted by each sensor during interval k, entering the freeway h intervals before k; time-varying travel time measures are considered.

The state variables $b_{ij}(k)$ are assumed to be stochastic in nature and evolve in some independent random walk process, as shown by the state equation:

$$b_{ij}(k+1) = b_{ij}(k) + w_{ij}(k) \tag{8}$$

for all feasible O-D pairs (i,j) where $w_{ij}(k)$s are independent Gaussian white noise sequences with zero mean and covariance matrix $\mathbf{Q}$.

The structural constraints should be satisfied for the state variables:

$$\begin{aligned} &b_{ij}(k) \geq 0 \qquad i = 1, \ldots, I \qquad j = 1, \ldots, J \\ &\sum_{j=1}^{J} b_{ij}(k) = 1 \qquad i = 1, \ldots, I \end{aligned} \tag{9}$$

where $b(k)$ is the column vector containing all feasible O-D pairs, ordered according to each entry ramp. Equality constraints whose sum is equal to one are imposed to ensure consistence with the definition of state variables in regard to proportions. When the measurement equations related to the state variables and observations are solved numerically, the satisfaction of these constraints is checked. This is not usually the case in implementations of the filter. In this case, the measurement equation becomes

$$z(k) = \begin{pmatrix} \mathbf{H}(\mathbf{k}) \\ \mathbf{E} \end{pmatrix} \mathbf{b}(\mathbf{k}) + \begin{pmatrix} v'(k) \\ 0 \end{pmatrix} \tag{10}$$

where $v'_{ij}(k)$s are independent Gaussian white noise sequences with zero mean and covariance matrix $\mathbf{R}'$, leading to a singular covariance matrix for the whole random noise vector:

$$\mathbf{V}[v(k)] = \mathbf{R} = \begin{bmatrix} \mathbf{R}' & 0 \\ 0 & 0 \end{bmatrix}$$

The size of matrix $\mathbf{R}$ is $(I + J + P)$.

Because the time-varying travel times have to be taken into account so that congestion can be modeled, time-varying delays from entries to sensor positions have to be considered (they are described in the building process of the observation equations); thus, on-ramp entry volumes for $M + 1$ intervals $k, k-1, \ldots, k-M$ must also be considered. State variables for intervals $k, k-1, \ldots, k-M$ are required for modeling interactions between time-varying O-D patterns, counts on sensors, and travel time delays from on-ramps to sensor positions.

Let $\mathbf{b(k)}$ be a column containing state variables for intervals k, $k-1, \ldots, k-M$ of dimension $(M+1) \times IJ$.

$$\mathbf{b}(\mathbf{k})^T = [b(k), b(k-1), \ldots, b(k-M)]^T \tag{11}$$

The state equations have to be written with a matrix operator D for shifting one interval [Chang and Wu (*8*)], which allows for eliminating the state variable for the last time interval (i.e., $k-M$) as

$$\mathbf{b}(\mathbf{k}+\mathbf{1}) = \mathbf{D}\mathbf{b}(\mathbf{k}) + \mathbf{w}(\mathbf{k}) \tag{12}$$

where $\mathbf{w(k)}^T = (w(k)\ 0 \cdots 0)$ is a white noise sequence with zero mean and singular covariance matrix:

$$\mathbf{V}[\mathbf{w}(\mathbf{k})] = \mathbf{W} = \begin{bmatrix} \mathbf{Q} & 0 \\ 0 & 0 \end{bmatrix}$$

and where $\mathbf{Q}$ of dimension IJ has been previously defined. It is usually a diagonal matrix. The following multinomial variance pattern has been successfully tested in the computational experiments:

$$\mathbf{D} = \begin{bmatrix} \mathbf{I}_{IJ} & 0 & \ldots & 0 \\ \mathbf{I}_{IJ} & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & \mathbf{I}_{IJ} & 0 \end{bmatrix}$$

The time-varying linear operator that relates O-D patterns and current observations for time interval k in Equation 4 is

$$\begin{pmatrix} \mathbf{H}(\mathbf{k}) \\ \mathbf{E} \end{pmatrix} = \begin{pmatrix} \mathbf{A}\mathbf{U}(\mathbf{k})^T \mathbf{F}(\mathbf{k}) \\ \mathbf{B} \quad 0 \quad \ldots \quad 0 \end{pmatrix} \tag{13}$$

where

- $\mathbf{E}$ = matrix of row dimension I containing 0 for columns related to state variables in time intervals $k-1, \ldots, k-M$ and $\mathbf{B}$ for time interval k,
- $\mathbf{B}$ = matrix of dimension $I \times IJ$ defining equality constraints (sum to 1 in O-D proportions for each entry) for state variables in time interval k,
- $\mathbf{F(k)}$ = matrix of dimensions $(1+M)IJ \times (1+M)IJ$ consisting of diagonal matrices $f(k), \ldots, f(k-M)$ containing input on-ramp volumes (applies to each O-D pair and time interval) and each $f(.)$ = squared diagonal matrix of dimension IJ,
- $\mathbf{g(k)}$ = column vector of O-D flows of equipped vehicles for time intervals $k, k-1, \ldots, k-M$,
- $U(k)$ = matrix of dimensions $(1+M)IJ \times (1+M)(J+P)$ consisting of diagonal matrices $U(k), \ldots, U(k-M)$ containing zeroes and ones, as defined above, and
- $\mathbf{A}$ = matrix of dimensions $(J+P) \times (1+M)(J+P)$ that adds up for a given sensor q (main section or off-ramp) traffic flows from any previous on-ramps arriving at the sensor at interval k, assuming their travel times are $t_{iq}(k)$.

Let

$$\mathbf{H}(\mathbf{k})\mathbf{b}(\mathbf{k}) = \mathbf{A}\mathbf{U}(\mathbf{k})^T \mathbf{F}(\mathbf{k})\mathbf{b}(\mathbf{k}) \approx \begin{pmatrix} s(k) \\ y(k) \end{pmatrix} \tag{14}$$

be a part of the observation equations, where the linear operator $\mathbf{H(k)}$ relates dynamic O-D proportions, dynamic travel time delays, and dynamic on-ramp entry flows to dynamic counts on sensors (main section and off-ramp) for equipped vehicles.

The space–state formulation is

$$z(k)=\begin{pmatrix}\mathbf{H(k)}\\ \mathbf{E}\end{pmatrix}\mathbf{b(k)}+\begin{pmatrix}v'(k)\\ 0\end{pmatrix}=\mathbf{R(k)b(k)}+v(k) \tag{15}$$

The recursive linear Kalman filter approach, well suited for online applications, has been implemented in MatLab with simulated data for the test site.

KF algorithm:

Let K be the total number of time intervals for estimation purposes and M = maximum number of time intervals for the longest trip.

Initialization:

$\mathbf{b}_k^k = \mathbf{b}(0)$ $K = 0$ build constant matrices and vectors: e, **A**, **B**, **D**, **E**, **R**, **W**, where each time interval and each row is set to the proportion $1/J_i$

$\mathbf{P}_k^k = \mathbf{V[b(0)]}$

Prediction step:

$\mathbf{b}_{k+1}^k = \mathbf{D}\mathbf{b}_k^k$

$\mathbf{P}_{k+1}^k = \mathbf{D}\mathbf{P}_k^k\mathbf{D}^T + \mathbf{W}$

Kalman gain computation:

Get observations of counts and travel times:

$q(k+1)$, $s(k+1)$, $y(k+1)$, $t_{ij}(k+1)$ $t_{ip}(k+1)$.

Build $z(k+1)$, $\mathbf{F(k+1)}$, $\mathbf{U(k+1)}$.

Build $\mathbf{R}_{k+1} = \mathbf{R(k+1)}$.

Compute $\mathbf{G}_{k+1} = \mathbf{P}_{k+1}^k\mathbf{R}_{k+1}^T(\mathbf{R}_{k+1}\mathbf{P}_{k+1}^k\mathbf{R}_{k+1}^T + \mathbf{R})^-$ [where $(.)^-$ denotes the pseudo inverse].

Filtering:

Compute $\mathbf{d}_{k+1} = \mathbf{G}_{k+1}\lfloor z(k+1) - \mathbf{R}_{k+1}\mathbf{b}_{k+1}^k \rfloor$ filter for state variables and errors $\boldsymbol{\epsilon}_{k+1} = \lfloor z(k+1) - \mathbf{R}_{k+1}\mathbf{b}_{k+1}^k \rfloor$.

Search maximum step length $0 \le \alpha \le 1$ such that $\mathbf{b}_{k+1}^{k+1} = \mathbf{b}_{k+1}^k + \alpha\,\mathbf{d}_{k+1} \ge 0$.

$\mathbf{P}_{k+1}^{k+1} = (\mathbf{I} - \mathbf{G}_{k+1}\mathbf{R}_{k+1})\mathbf{P}_{k+1}^k$

Iteration:

$k = k+1$

if $k = K$, EXIT; otherwise GO TO Prediction Step.

Exit:

Print results.

PRELIMINARY RESULTS

Table 3 presents a sample of the results of applying the Kalman filter approach for 5-min travel time forecasting. The computational results were obtained by using the variances of the 5-min samples for Tuesdays stored in the historical database and the real-time travel time measurements for a specific Tuesday. A quantitative estimation of the quality of the prediction is given by the correlation coefficient between the two series, $R^2 = 0.9863$, and a mean absolute relative error of 0.0354. Taking into account that both series of data—measured and predicted travel times—are time series, a measure of the fitting quality can be defined in relation to Theil's coefficients (*15*).

Theil's inequality coefficient is a measure of how close two time series are (overcoming the effect of outliers in root-mean-square estimators) and is given by

$$U = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\hat{y}_i^2} + \sqrt{\frac{1}{n}\sum_{i=1}^{n}y_i^2}} \tag{16}$$

Bounded between 0 and 1, $U = 0$ can be interpreted as a perfect fitting between the two series; $U = 1$ represents an unacceptable discrepancy. Values of $U > 0.2$ recommend a rejection of the predicted series. In this case the value of U is $U = 0.02415735$, indicating that the matching is very good. However, Theil's coefficient can be decomposed in the three coefficients:

$$U_M = \frac{n(\bar{\hat{y}} - \bar{y})^2}{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2},\quad U_S = \frac{n(\hat{\sigma} - \sigma)^2}{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2},$$

$$U_C = \frac{2(1-\rho)n\,\hat{\sigma}\,\sigma}{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2},\quad U_M + U_S + U_C = 1 \tag{17}$$

where $\bar{\hat{y}}$ and $\bar{y}$ are, respectively, the means of the measurements and predictions, $\hat{\sigma}$ and σ are the standard deviations, and ρ is the correlation coefficient. U_M, the bias proportion, can be considered as a measure of the systematic error; U_S, the variance proportion, identifies the predicted series' ability to reproduce the variability of the observed time series; and U_C, the covariance proportion, is a measure of the nonsystematic error. In the present case the corresponding values are $U_M = 0.088641663$, $U_S = 0.002415572$, and $U_C = 0.913031427$. The small values of U_M and U_S certify the quality of the prediction.

O-D ESTIMATION

A set of computational experiments has been conducted with simulated data, assuming in all cases a fixed 30% rate of equipped vehicles. O-D pattern initialization is noninformative (every off-ramp from

TABLE 3 Sequence of Computations in Travel Time Forecasting Algorithm

k	Measured Travel Time	$R(k)$	$A(k)$	State est. = $A(k-1) * x(k-1)$	$P(t)-$	Kalman Gain	$P(t)+$	Predict.
0	325	7,711	1	320	1,000.0000	0.1147	885.202617	320.5739
1	306.4375	1,600.7461	0.9428	320.5739	885.2026	0.3560	569.997534	315.5402
2	359.5789	35,845.2964	1.1734	297.5180	506.7457	0.0139	499.681694	298.3831
3	314.0588	1,199.9377	0.8734	350.1277	688.0151	0.3644	437.285992	336.9833
4	332.4	2,419.1733	1.0584	294.3237	333.5793	0.1211	293.156075	298.9378
5	316.7778	1,386.0617	0.9530	316.3959	328.3968	0.1915	265.493911	316.4690

TABLE 4 O-D Pairs 1-1 to 1-10: Convergence to Truly O-D Proportion for Constant O-D Pattern

	1-1	1-2	1-3	1-4	1-5	1-6	1-7	1-8	1-9	1-10
O-D Pairs 1-1 to 1-10—Fixed O-D Pattern—Static O-D Flows										
RMSE $\times 10^{-2}$										
Interval length 90 s										
Uncongested	6.42	3.29	6.23	4.80	5.11	6.07	5.19	4.82	2.86	6.03
Congested	6.46	3.32	6.24	4.83	5.20	5.97	5.23	5.50	2.75	6.11
O-D Pairs—Fixed O-D Pattern—Time Sliced O-D Flows										
4th time slice: RMSE $\times 10^{-2}$										
Interval length 90 s										
Time sliced O-D flows	6.35	3.27	6.14	4.75	4.94	5.92	5.35	5.20	3.69	6.13

NOTE: Sets 1 and 2 of computational experiments—time horizon of 1 h, RMSE values multiplied by 10^{-2}.

one on-ramp has the same probability) for the 74 O-D pairs in the site's model, a hard initialization. In the first set, two fixed O-D patterns with static O-D flows have been used for testing purposes, with a time horizon of 1 h. There is an O-D pattern for uncongested conditions and another pattern for congested conditions. The test shows that the proposed Kalman filtering approach converges successfully to the true results in a few iterations. The second set of computational experiments has been conducted with time-sliced O-D flows, totalizing the same amount of demand as in the first test set, but with the time horizon split into four time intervals of 15 min each; the demand is also distributed accordingly to account for the 15%, 25%, 35%, and 25% of the total demand in each interval. That means that although the O-D pattern (that is, the proportions) is fixed, the O-D flows are time dependent. That is, the O-D pattern is fixed but not the O-D flows, which are time dependent. The results can be summarized as follows: for time intervals in which traffic flow varies from free flow to dense, but not yet saturation conditions, the filtering approach works as expected and its performance seems unaffected as traffic flows become congested.

Table 4 summarizes the values of the root-mean-square error (RMSE) for each O-D proportion for both sets of experiments. It compares the RMSE error values for congested and uncongested tests for some O-D pairs (RMSE on O-D proportions) at the last iteration. The results show no significant differences in the accuracy of the estimates of target O-D proportions. Initialization of covariance has a key effect on convergence in accordance with the experience reported by other researchers. The convergence and RMSE are equivalent in both cases for the same set of O-D pairs.

Figure 3 illustrates a couple of additional cases for other path flows from entry to exit ramps that correspond to shorter distances (Entry 4 to Exit 9 and Entry 10 to Exit 11). The graphics show how, in both cases, the filter algorithm converges to the true values of the O-D proportions in the synthetic simulation experiment. These values are 0.2411 for Pair (5-11) and 0.313 for Pair (10-11).

CONCLUSIONS AND FUTURE RESEARCH

Bluetooth sensors that detect mobile devices have proved to be a mature technology that provides sound measurements of average speeds and travel times between sensor locations. This paper has developed and tested a Kalman filter approach for travel time forecasting based on these measurements. The result proves the quality of the forecasts.

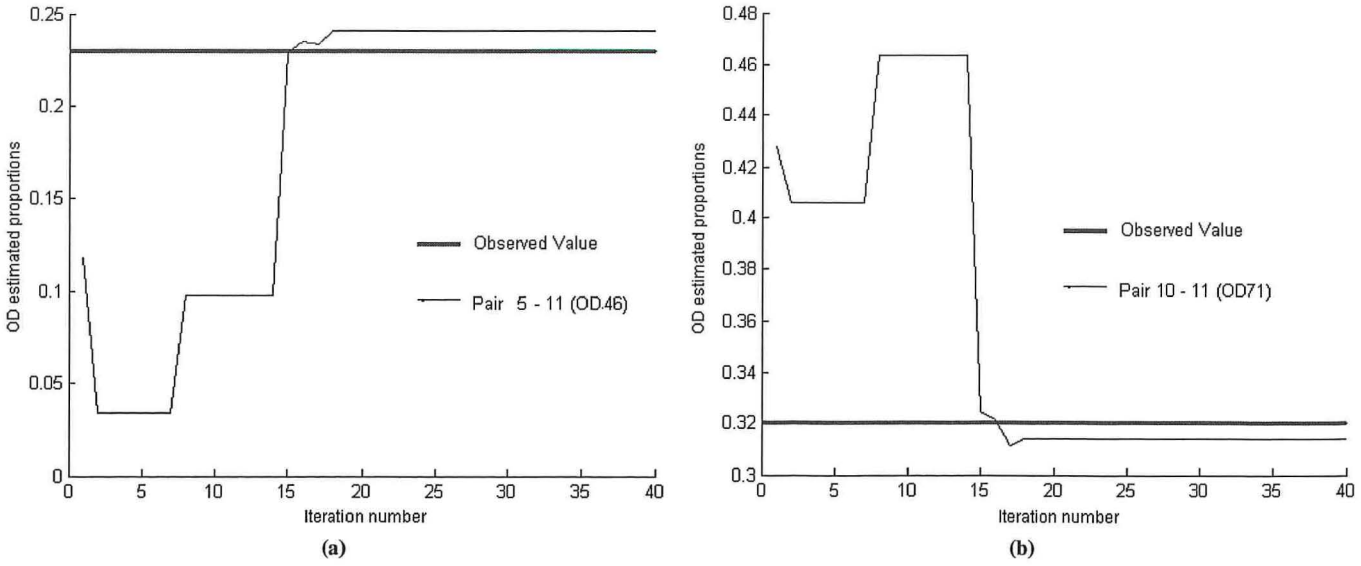


FIGURE 3 Convergence for (*a*) O-D Pairs 5–11 and (*b*) O-D Pairs 10–11.

How data available from this technology can be used to estimate dynamic origin-to-destination matrices in motorways has also been explored by proposing an ad hoc linear Kalman filter approach. Results of the conducted experiments with simulation data prove that the approach works well for uncongested and congested conditions, but properly tuning the initialization matrices is critical in both situations.

Because precision of the estimated O-D pattern is also affected by interval length, an adaptive time-varying scheme for time interval length, according to congestion, should also be included in the research in the near future.

ACKNOWLEDGMENTS

The authors acknowledge the firms Abertis, operator of the AP-7 Motorway, and Bitcarrier, developer of the Bluetooth sensors, for their kindness in making the data used in this paper available; and Pilar Muñoz, of the Department of Statistics and Operations Research at the Technical University of Catalonia, for her kind advice on Kalman filtering. The research undertaken in this project has been funded by projects SIMETRIA and MARTA of the Spanish R+D National Program.

REFERENCES

1. Waller, S. T., Y.-C. Chiu, N. Ruiz-Juri, A. Unikrishnan, and B. Bustillos. *Short Term Travel Time Prediction on Freeways in Conjunction with Detector Coverage Analysis.* Technical Report FHWA/TX-08/0-5141-1. Center for Transportation Research, University of Texas at Austin for Texas Department of Transportation, Austin, 2007.
2. Sanfeliu, A., A. Punsola, Y. Yoshimura, M. R. Llácer, and M. D. Gramunt. Legal Challenges for Networking Robots Deployment in European Urban Areas: The Privacy Issue. Workshop on Network Robot Systems. Presented at IEEE International Conference on Robotics and Automation, Kobe, Japan, 2009.
3. Nanthawichit, C., T. Nakatsuji, and H. Suzuki. Application of Probe-Vehicle Data for Real-Time Traffic-State Estimation and Short-Term Travel-Time Prediction on a Freeway. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1855,* Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 49–59.
4. Kuchipudi, C., and S. I. J. Chien. Development of a Hybrid Model for Dynamic Travel Time Prediction. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1855,* Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 22–31.
5. Cheng, P., Q. Zhijun, and B. Ran. Particle Filter-Based Traffic State Estimation Using Cell Phone Network Data. *Proc., Intelligent Transportation Systems Conference,* Toronto, Ontario, Canada, 2006.
6. Kalman, R. E. A New Approach to Linear Filtering and Prediction Problem. *Journal of Basic Engineering,* Vol. 82, No. 1, 1960, pp. 35–45.
7. Van Der Zijpp, N. J., and R. Hamerslag. Improved Kalman Filtering Approach for Estimating Origin–Destination Matrices for Freeway Corridors. In *Transportation Research Record 1443,* TRB, National Research Council, Washington, D.C., 1994, pp. 54–64.
8. Chang, G. L., and J. Wu. Recursive Estimation of Time-Varying Origin–Destination Flows from Traffic Counts in Freeway Corridors. *Transportation Research B,* Vol. 28, No. 2, 1994, pp. 141–160.
9. Chu, L., J.-S. Oh, and W. Recker. Adaptive Kalman-Filter-Based Freeway Travel Time Estimation. Presented at 84th Annual Meeting of the Transportation Research Board, Washington, D.C., 2005.
10. Work, D., O.-P. Toossavainen, S. Blandin, A. Bayen, T. Iwuchukwu, and T. Tracton. An Ensemble Kalman Filtering Approach to Highway Traffic Estimation Using GPS Enabled Mobile Devices. *Proc., 47th IEEE Conference on Decision and Control,* Cancun, Mexico, Dec. 9–11, 2008, pp. 5062–5068.
11. Bell, M. G. H. The Real Time Estimation of Origin–Destination Flows in the Presence of Platoon Dispersion. *Transportation Research Part B,* Vol. 25, No. 2–3, 1991, pp. 115–125.
12. Codina, E., and J. Barceló. Adjustment of O-D Matrices from Observed Volumes: An Algorithmic Approach Based on Conjugate Gradients. *European Journal of Operations Research,* Vol. 155, 2004, pp. 535–557.
13. Ben-Akiva, M., M. Bierlaire, D. Burton, H. N. Kotsopoulos, and R. Mishalani. Network State Estimation and Prediction for Real-Time Traffic Management. *Networks and Spatial Economics,* Vol. 1, 2001, pp. 293–318.
14. Mahmassani, H. S., and X. Zhou. Transportation System Intelligence: Performance Measurement and Real-Time Traffic Estimation and Prediction in a Day-to-Day Learning Framework. In *Advances in Control, Communication Networks, and Transportation Systems* (E. Abed, ed.), Birkhauser Boston, Mass., 2005.
15. Theil, H. *Applied Economic Forecasting.* North-Holland, Netherlands, 1966.

The Transportation Demand Forecasting Committee peer-reviewed this paper.

Real-Time Short-Term Traffic Speed Level Forecasting and Uncertainty Quantification Using Layered Kalman Filters

Jianhua Guo and Billy M. Williams

Short-term traffic condition forecasting has long been argued as essential for developing proactive traffic control systems that could alleviate the growing congestion in the United States. In this field, short-term traffic condition level forecasting and short-term traffic condition uncertainty forecasting play an equally important role. Past literature showed that linear stochastic time series models are promising in modeling and hence forecasting traffic condition levels and traffic conditional variance with workable performance. On the basis of this finding, an autoregressive moving average plus generalized autoregressive conditional heteroscedasticity structure was proposed for modeling the station-by-station traffic speed series. An online algorithm based on layered Kalman filter was developed for processing this structure in real time. Empirical results based on real-world station-by-station traffic speed data showed that the proposed online algorithm can generate workable short-term traffic speed level forecasts and associated prediction confidence intervals. Future work is recommended to develop and test a proactive traffic control system in a simulated environment, to refine the uncertainty modeling through a stochastic volatility model, and to extend uncertainty modeling and forecasting to link level and network level.

Congestion has been a growing concern for operating the surface transportation systems in the United States. Targeting this issue, many solutions have been put forward, collectively known as intelligent transportation systems. In this direction, proactive traffic control systems are promising in combating or mitigating the negative effects of congestion by using predicted traffic conditions in these systems. Compared with reactive traffic control systems that react at most to current traffic conditions, the proactive control systems might have a chance of eliminating or postponing the onset of congestion given accurate anticipation of the traffic condition in the near future. For this purpose, short-term traffic condition forecasting has been extensively investigated in the past decades (*1–3*).

Short-term traffic condition forecasting includes the forecasting of the traffic condition level and the forecasting of traffic condition uncertainty. In short-term traffic condition level forecasting, traffic conditions are in general represented in relation to conventional traffic variables (e.g., flow rate or speed), which are readily obtainable through the widely deployed loop detectors along many highway systems. Compared with flow rate, which in general represents the demand of transportation systems, speed is more directly related to the roadway operation status represented by occupancy; in other words, one flow rate corresponds to two possible occupancies whereas one speed corresponds only to one (*4*).

In addition to level forecasting, traffic condition uncertainty forecasting, usually in regard to traffic condition variance, is gradually gaining attention in the transportation community (*5*). In the context of short-term traffic condition forecasting, an important issue is to generate the reliability of the forecast traffic condition levels so that prediction confidence levels can be constructed for supporting the development of the proactive traffic control systems.

Therefore, there is an ongoing need for a short-term traffic condition forecasting algorithm with the ability of generating level and uncertainty forecasts. In this paper an online algorithm based on a layered Kalman filter structure will be presented for predicting station-by-station traffic speed data. This algorithm has the ability of generating the level predictions and the related prediction confidence intervals in real time. Following a literature review, the data and methodology are described; afterward, the empirical results are presented for demonstrating the abilities of the proposed algorithm.

LITERATURE REVIEW

As mentioned in the introduction section, many methods have been proposed for short-term traffic condition forecasting. However, they are focused mainly on traffic level forecasting, and very few studies are conducted on forecasting traffic condition uncertainty. In this section, the methodological features of short-term traffic condition forecasting methods are highlighted.

Level Forecasting

Traffic condition level forecasting has been investigated intensively in the past decades. The methods can generally be divided into the following categories: the heuristic method, linear method, nonlinear method, hybrid method, and traffic flow theory-based method (*3*).

The heuristic methods were first developed in the early 1980s, including random walk, historical average, informed historical average, and urban traffic control system predictors. These methods are easy to implement; human expertise and engineering judgment are necessary and critical for a successful field deployment (*1*).

Linear methods were developed since the late 1970s. The methods were developed on the basis of the assumption that traffic condition

J. Guo, Federal Highway Administration, 6300 Georgetown Pike, McLean, VA 22101. B. M. Williams, Department of Civil, Construction, and Environmental Engineering, North Carolina State University, Campus Box 7908, Raleigh, NC 27695. Corresponding author: B. M. Williams, billy_williams@ncsu.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2175, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 28–37.
DOI: 10.3141/2175-04

data are linear stochastic in nature so that a linear structure can be applied to capture and hence forecast traffic condition. Typical methods include the seasonal autoregressive integrated moving average (ARIMA) (*1, 6*), nonseasonal ARIMA models (*7, 8*), Holt–Winter method (*9*), spectral representation (*10*), and multivariate time series models (*11, 12*). Filtering approaches, for example, Kalman filter (*3, 13–17*), also belong to this category in that linear operations are inherently dominant in their structure.

Unlike linear methods, the nonlinear methods assume that traffic condition data are nonlinear in nature and nonlinear structure can be applied to capture and hence predict traffic condition. Typical methods include neural networks (*18–21*) and *k*-nearest neighbor (*22, 23*). These methods have the desirable attribute of adapting to changing traffic condition; however, expanding the historical database needed for the adaptation decreases their computational efficiency over time, which is critical for short-term traffic condition forecasting. By comparing linear and nonlinear approaches, Smith et al. suggested that a linear stochastic model is more appropriate for modeling traffic condition data (*24*).

Hybrid methods could be the combination of multiple methods or the combination of predictions from multiple methods (*25, 26*). Generally these methods have complex structures that are not desirable for wide field implementation.

Traffic flow theory–based approaches directly use traffic state propagation based on traffic flow theory for traffic condition prediction (*27*). In general, this method can produce desirable forecasting accuracy; usually, however, it will require correct traffic state separation and the forecasting horizon is in general limited to a very short time period, for example, 20 s.

Uncertainty Forecasting

Compared with the extensive investigations in traffic condition level forecasting, traffic condition uncertainty forecasting lacks adequate attention. Originally, very few methods were used in this direction; examples include the bootstrap method (*28, 29*) and the method based on unconditional forecasting error variance (*14*). However, the reported empirical evidence shows that all these methods either lack reasonable coverage (ideally 95% coverage for 5% significant level interval) or the intervals cannot vary with time.

Most recently, Guo (*3*) and Guo et al. (*17*) applied a generalized autoregressive conditional heteroscedasticity (GARCH) model for estimating the time-varying conditional variance of traffic demand series; in addition, an online algorithm based on Kalman filter was designed to process the GARCH model in real time. Empirical results showed that the proposed algorithm can generate workable prediction confidence intervals in relation to kickoff percentage and the prediction confidence interval width-to-level ratio. Kamarianakis et al. (*30*) applied the GARCH model in modeling the relative velocity (defined as volume divided by occupancy), and Sohn and Kim (*31*) applied the GARCH model in forecasting link travel time variability; however, no online algorithm was proposed in these studies for processing the model in real time, which is critical for online applications. Tsekeris and Stathopoulos proposed to use ARIMA + GARCH and autoregressive fractionally integrated moving average + fractionally integrated asymmetric power autoregressive conditional heteroscedasticity (ARCH) (ARFIMA + FIAPARCH) for forecasting traffic volatility in real time for urban networks (*32*). Their online approach, however, was carried out recursively with the quasi-maximum likelihood estimation method provided in the G@RCH software, which will, on one hand, incur high computational demand and the possibility of estimation error due to nonconvergence and, on the other hand, increase the complexity for field implementation.

In summary, although short-term traffic condition level forecasting has been widely investigated in the past decades, investigation into the short-term traffic condition uncertainty forecasting is less developed. According to the literature review above, the linear method is promising for short-term traffic condition level forecasting, and the GARCH model has been shown to have the desirable attribute of capturing the time-varying conditional variance. Therefore, in this study, the time series model approach, that is, an autoregressive moving average (ARMA) + GARCH structure, is explored for modeling and predicting short-term traffic speed series.

METHODOLOGY

In this section, data used in this study are described, and the ARMA + GARCH model is presented, based on which a layered Kalman filter structure is designed for predicting speed level and conditional speed series variance.

Data

Traffic speed data collected for 10 stations along I-80 in the Bay Area of California were used in this study through the PeMS system. The data collection time period is from May 4, 2006, to July 3, 2006, with the data collection time interval as 5 min. The California data had passed quality screening tests before their use in this work. The overview of the speed data for these stations is presented in Table 1. These traffic speed data are also described by two categories: speed >30 mph (uncongested category) and speed <30 mph (congested category).

Using speed data from Station 400329 as a typical example, Figure 1 shows the speed pattern across a whole week. From Figure 1, it can be seen that for the uncongested category, the speed observations are usually oscillating around a high speed level with occasional speed drops. Some of the speed drops will not degenerate the series into the congested category, but some will degenerate the traffic into congestion before traffic comes back to the uncongested category. For the uncongested traffic category, the stable speed pattern implies a high predictability; an adaptive mechanism is needed for switching the forecasting from the uncongested category into the congested category.

Speed Series Dynamics

Determining the speed series dynamics is necessary for constructing an online forecasting algorithm. In this section, on the basis of the discussion of the linear stochastic nature of traffic condition, a linear time series model is used to model the evolution of the speed series. This model structure includes an ARMA component for conditional level modeling and a GARCH component for conditional variance modeling. Note that the conditional variance modeling or the heteroscedasticity modeling was first proposed in Engle (*33*) as ARCH, and then extended in Bollerslev (*34*) to GARCH, for modeling the volatility in financial series.

TABLE 1 Data Overview

Station	Number of Lanes	Mile Marker	Speed < 30 mph		Speed > 30 mph		Total Length
			Count	Percentage	Count	Percentage	
401052	4	13.41	512	2.91	17,056	97.09	17,568
400329	4	13.79	409	2.33	17,159	97.67	17,568
401195	4	14.47	616	3.51	16,952	96.49	17,568
400445	4	15.97	453	2.58	17,115	97.42	17,568
400443	4	16.32	285	1.62	17,283	98.38	17,568
401209	4	19.29	214	1.22	17,354	98.78	17,568
400976	4	19.92	265	1.51	17,303	98.49	17,568
400838	4	20.24	482	2.74	17,086	97.26	17,568
400430	4	20.64	662	3.77	16,906	96.23	17,568
400865	4	20.96	376	2.14	17,192	97.86	17,568

NOTE: Missing values were imputed in the PeMS system.

Given a discrete speed time series $\{X_t\}$, the ARMA + GARCH structure is defined as

$$\begin{aligned}(1-\phi B)X_t &= (1-\theta B)\epsilon_t \\ \epsilon_t &= \sqrt{h_t}e_t \\ h_t &= \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \beta_1 h_{t-1} \\ e_t &\sim \mathrm{IN}(0,1)\end{aligned} \tag{1}$$

where

B = backshift operator such that $BX = X_{t-1}$,
ϕ = autoregressive parameter,
θ = moving average parameter,
h_t = conditional variance at time t, that is, $\epsilon_t|\Psi_{t-1} \sim N(0, h_t)$ with Ψ_{t-1} as the information up to time $t-1$,
α_0 = positive constant coefficient,
α_1 = nonnegative coefficients of lagged sample variance ϵ_{t-1}^2,
β_1 = nonnegative coefficients of lagged conditional variance h_{t-1}, and
IN = independent normal.

In the model above, the order of both ARMA and GARCH was selected as (1, 1). This parsimonious model order selection will facilitate the online algorithm development as well as meet the requirement of capturing speed series dynamics; in addition, the online model based on ARMA + GARCH is expected to compensate for unexpected speed dynamic shifts that might be happening in the field.

The ARMA component is a localized stationary model without considering the seasonality of typical traffic conditions, that is, weekly pattern in demand series used in, for example, Williams (*1*) and Williams and Hoel (*6*). This selection is in accordance with the typical speed pattern as shown in Figure 1, in which most of the traffic speed data fall into the stable uncongested category; in addition, the online algorithm that will be developed on the basis of the ARMA model is expected to handle any less significant nonstationarity in the traffic speed series.

Kalman Filter Design

The proposed layered Kalman filter structure includes two Kalman filters, and the design of the layered Kalman filter follows two steps,

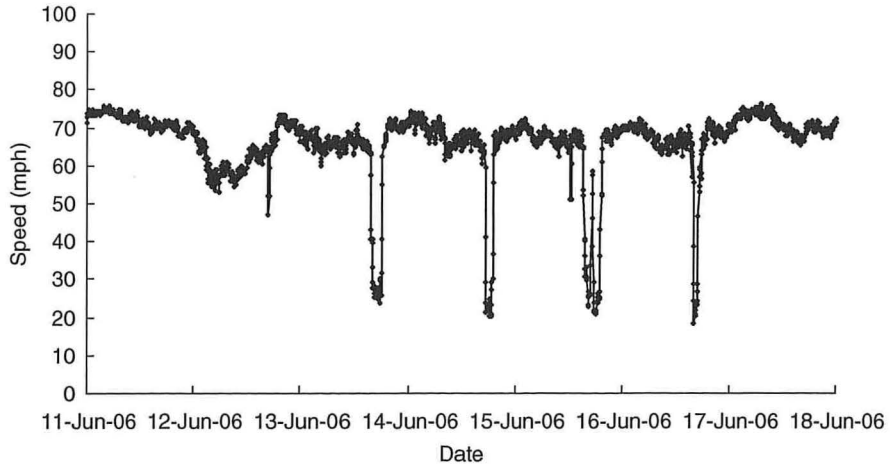


FIGURE 1 Typical speed series pattern: Station 400329.

that is, Step 1, the Kalman filter design based on the ARMA (1, 1) model, and Step 2, the Kalman filter design based on the GARCH (1, 1) model. In Step 1, the ARMA (1, 1) model can be reorganized as

$$x_t = \phi x_{t-1} + \theta e_{t-1} + e_t \quad (2)$$

where

x_t = time series variable,
t = time index, and
e_t = white noise process with mean zero and variance σ_e^2.

Considering the state space approach framework, further generalization can be reached by defining a state transition equation as

$$w_t = \Phi w_{t-1} + a_t \quad (3)$$

and an observation equation as

$$Y_t = X_t^T w_t + e_t \quad (4)$$

where

e_t = white noise process with mean zero and variance σ_e^2,
a_t = white noise process with mean zero and variance σ_a^2,
T = matrix operator of transpose,
$w_t = (\phi\ \theta)^T$ = state variable,
Φ = diag$\{\lambda^{-1/2}\}$ = state transition matrix, with λ defined as a forgetting factor,
Cov$(a_t a_t^T) = Q_t$ = state noise covariance matrix,
$Y_t = x_t$ = current observation,
$X_t = (x_{t-1}\ e_{t-1})^T$ = time varying observation matrix, and
Cov$(v_t v_t^T) = R_t$ = observation noise covariance matrix.

The combination of Equations 3 and 4 formulates a Kalman filter, in which the ARMA model parameters are treated as hidden state and an instrument of random walk with forgetting factor was used for its evolution, and the speed series is treated as the driving observation process. This Kalman filter can be readily solved by using standard Kalman recursion equations (*35*).

In Step 2, the GARCH (1,1) is first presented by Bollerslev (*34*) as

$$\epsilon_t^2 = \alpha_0 + (\alpha_1 + \beta_1)\epsilon_{t-1}^2 - \beta_1 \eta_{t-1} + \eta_t \quad (5)$$

where

ϵ_t = defined as in Equation 1,
η_t = serially uncorrelated with mean zero, and
t = time index.

A reparameterization yields

$$\epsilon_t^2 = \alpha_0 + \alpha\epsilon_{t-1}^2 + \beta\eta_{t-1} + \eta_t \quad (6)$$

with $\beta = -\beta_1$ and $\alpha = \alpha_1 + \beta_1$. Then the state space model is defined as

Observation equation:

$$\epsilon^2 = \begin{pmatrix} 1 & \epsilon_{t-1}^2 & \eta_{t-1} \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha \\ \beta \end{pmatrix} + \eta_t \quad (7)$$

State transition equation:

$$\begin{pmatrix} \alpha_0 \\ \alpha \\ \beta \end{pmatrix}_t = \text{diag}\{\lambda^{-1/2}\} \begin{pmatrix} \alpha_0 \\ \alpha \\ \beta \end{pmatrix}_{t-1} + a_t \quad (8)$$

where Cov$(\eta_t \eta_t^T)$ equals R_t: observation noise covariance matrix and Cov$(a_t a_t^T)$ equals Q_t: system noise covariance matrix.

The combination of Equations 7 and 8 formulates a Kalman filter, in which the reparametrized parameter vector is treated as the hidden state and the squared residual series from the ARMA component is treated as the driving observation process. Similarly this state–space model can be readily solved with the standard Kalman filter recursion.

EMPIRICAL RESULTS

In this section, first the ARMA + GARCH structure is validated; then the aggregated and disaggregated performance of the online algorithm (called "online") are presented and compared with those of the batch processing (called "batch") of the ARMA + GARCH model.

ARMA + GARCH Validation

The purpose of this section is to validate the ARMA + GARCH structure, in particular, the validation of the necessity of appending the GARCH component. For this purpose, first Table 2 presents the autocorrelations in the residuals after the ARMA (1, 1) model is applied to each station series. It is clear that the autocorrelations among the residuals are trivial up to Lag 24 for all stations. This demonstrates that the ARMA (1, 1) is adequate in capturing the autocorrelation structure in traffic speed series.

Even though the autocorrelations of the ARMA (1, 1) residual series are insignificant, the test of the autocorrelation structure in the squared residual series showed that this residual series has a changing conditional variance phenomenon called "heteroscedasticity" in the literature. In doing so, the heteroscedasticity test, that is, the conventional portmanteau Lagrange Multiplier test, is applied. Considering the inflation effects due to the length of the series, the test was broken into test by week and test by day. For each test, the whole residual series was partitioned into test units (week or day), and test results were reported in Table 3. It is clear from Table 3 that significant heteroscedasticity exists in the speed series across all stations, supporting the appending of the GARCH component.

Aggregated Performance

In this section, the aggregated performance of the proposed online algorithm and the batch processing was presented and compared. Because the proposed online algorithm is able to produce speed level forecasts and forecast confidence intervals, two categories of aggregated performance measures are used in this study, measures of forecasting accuracy and measures of prediction confidence interval validity. In addition, considering the need for the online algorithm to converge to its equilibrium state, the observations in the first 3 days are excluded from computing the performance measures.

TABLE 2 Autocorrelations of Residuals

Station	Up to Lag	Autocorrelation					
401052	6	0.020	**0.082**	0.023	0.011	0.006	−0.014
	12	−0.027	−0.014	0.005	−0.006	−0.010	−0.005
	18	0.006	0.010	0.004	−0.016	0.013	0.019
	24	0.004	0.006	−0.010	−0.006	−0.007	−0.007
400329	6	0.005	0.057	**0.065**	−0.020	−0.011	0.002
	12	−0.007	0.020	−0.005	−0.005	0.022	−0.001
	18	0.003	**0.049**	−0.030	−0.010	0.012	−0.027
	24	0.000	−0.018	−0.027	−0.004	−0.014	−0.020
401195	6	0.002	0.024	**0.059**	−0.027	0.018	0.015
	12	0.018	0.024	−0.027	0.018	0.034	0.015
	18	0.008	0.018	−0.014	0.003	0.022	−0.031
	24	0.004	−0.007	−0.013	0.003	−0.008	0.001
400445	6	0.005	**0.045**	0.029	0.026	0.022	0.045
	12	−0.005	0.014	−0.008	0.008	0.043	0.034
	18	0.014	0.014	0.012	−0.014	0.034	−0.008
	24	−0.001	−0.038	0.002	−0.035	0.009	−0.036
400443	6	0.002	0.025	0.022	0.035	0.021	0.030
	12	0.022	0.010	0.001	0.010	0.030	**0.036**
	18	0.018	0.017	0.011	0.003	0.007	0.011
	24	0.003	−0.009	−0.021	−0.029	−0.006	−0.007
401209	6	−0.001	−0.017	0.011	0.011	0.007	−0.010
	12	0.011	−0.005	0.003	0.013	**0.024**	0.015
	18	−0.016	0.012	−0.018	−0.01	−0.009	0.001
	24	0.003	−0.008	0.014	−0.008	0.005	0.036
400976	6	0.008	**0.078**	0.058	0.035	0.039	0.031
	12	0.028	0.007	0.010	−0.037	−0.001	−0.026
	18	0.000	0.006	−0.035	−0.024	−0.026	0.000
	24	−0.023	−0.020	−0.008	−0.012	−0.028	0.007
400838	6	0.009	**0.083**	0.047	0.007	0.063	0.035
	12	0.000	0.009	0.031	−0.039	−0.007	0.006
	18	−0.014	0.021	−0.021	−0.005	−0.008	−0.004
	24	−0.006	−0.022	−0.004	−0.003	−0.012	−0.013
400430	6	0.006	**0.069**	0.053	0.013	0.049	0.027
	12	0.028	0.013	0.030	−0.018	0.017	−0.012
	18	−0.024	0.034	−0.010	−0.007	0.007	−0.002
	24	−0.004	−0.014	−0.007	−0.008	−0.016	−0.010
400865	6	0.002	**0.062**	0.037	0.026	0.044	0.026
	12	0.035	0.041	0.025	0.010	0.012	0.005
	18	−0.010	0.012	0.002	0.004	0.006	0.006
	24	−0.002	0.005	−0.020	−0.013	−0.011	−0.007

NOTE: The autocorrelations were computed up to Lag 24; bold values are the maximum autocorrelations for each station.

TABLE 3 Heteroscedasticity Testing Result

	By Week		By Date		
Station	Significant Unit	Total Unit	Significant Unit	Total Unit	Percentage
401052	8	10	36	61	59.02
400329	10	10	44	61	72.13
401195	10	10	36	61	59.02
400445	10	10	46	61	75.41
400443	10	10	45	61	73.77
401209	10	10	45	61	73.77
400976	10	10	42	61	68.85
400838	10	10	46	61	75.41
400430	10	10	44	61	72.13
400865	10	10	48	61	78.69

NOTE: Test by week is more powerful than test by date because of the inflation effect due to longer residual series. The portmanteau Lagrange Multiplier test was performed up to Lag 12, and the test result was considered as significant only when the maximum p-value for all 12 lags is less than .05.

Prediction Accuracy Performance

Three measures of forecasting accuracy were used in this study. Given X_t as the real observations, $\hat{X}_t$ as the forecasts, and n as the total number of observations processed, these measures are defined as follows:

Mean absolute error (MAE):

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}\left|X_t - \hat{X}_t\right| \tag{9}$$

Mean absolute percentage error (MAPE):

$$\text{MAPE} = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{X_t - \hat{X}_t}{X_t}\right| \tag{10}$$

Root-mean-square error (RMSE):

$$\text{RMSE} = \frac{1}{n}\sqrt{n\sum_{t=1}^{n}\left(X_t - \hat{X}_t\right)^2} \tag{11}$$

The prediction accuracy performance of the online algorithm and batch mode processing is presented in Figure 2, with each plot corresponding to one performance measure. Considering the typical speed pattern, the performance was computed for two groups, that is, speed >30 mph and speed <30 mph.

Multiple observations can be drawn from Figure 2. First, for speed >30 mph, the performance of the online algorithm and the batch processing are almost indistinguishable across all forecasting accuracy performance measures and all stations. That is not surprising in that for higher speed, the speed evolution is typically stable, indicating a high predictability for both approaches.

Second, for speed <30 mph, the online algorithm outperformed the batch model. On reflection, even though the batch mode processing used all speed data in the series for making the predictions, the speed drops as shown in typical speed patterns make it hard to generate a model that is optimal for both uncongested and congested groups. In

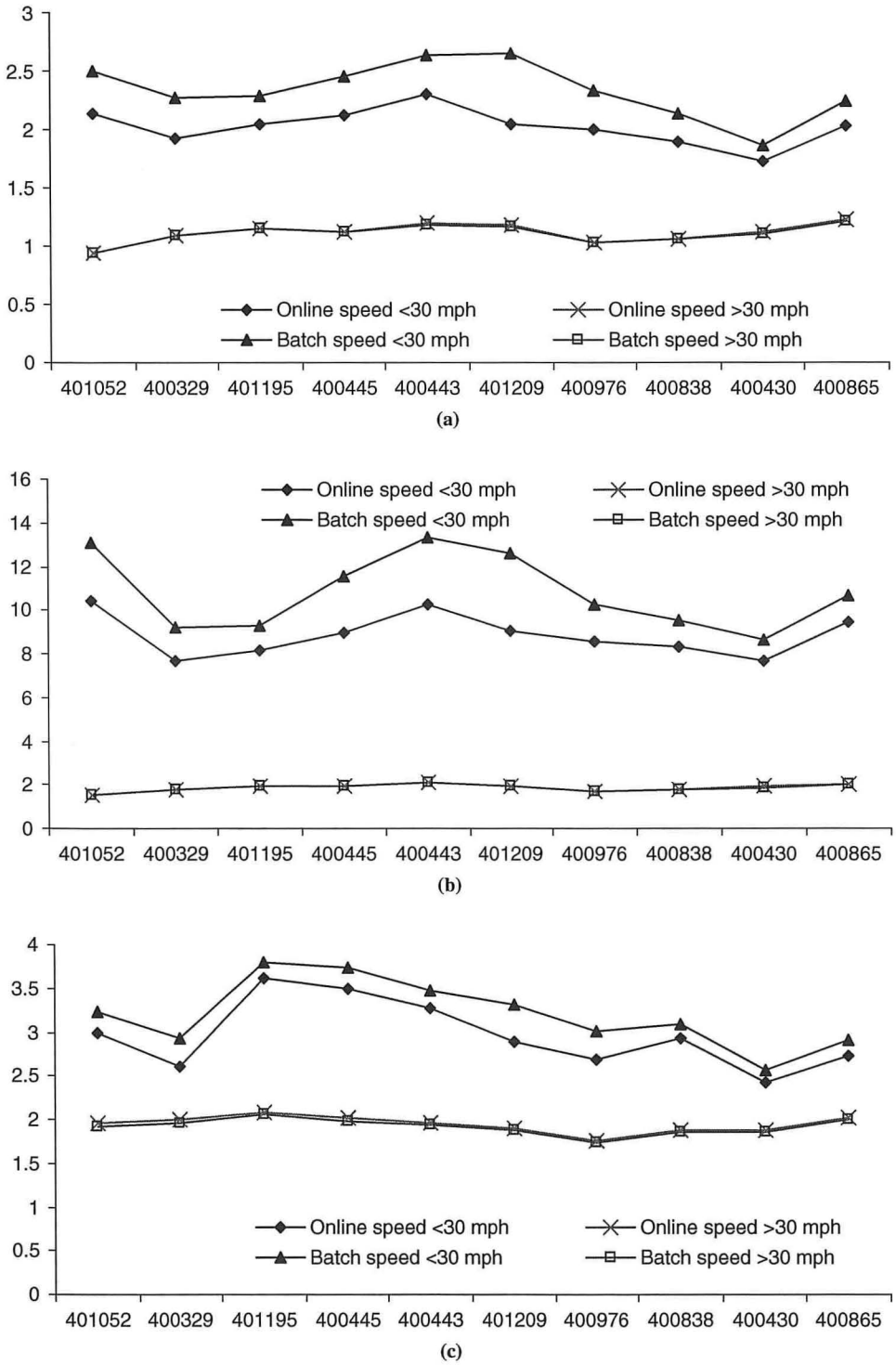


FIGURE 2 Forecasting accuracy performance: (*a*) MAE, (*b*) MAPE, and (*c*) RMSE (*x*-axis is station and *y*-axis is performance measure).

contrast, although only observations up to current time were used for the online algorithm, the real-time adjustment of the online algorithm structure enables the algorithm to be flexible and adaptive to the changing speed condition so that better forecasting accuracy can be achieved. That is desirable in that the online algorithm can meet the expectation of adapting to the changing traffic speed patterns in real time.

Finally, the online algorithm and batch processing can generate workable accuracy measures, for example, for congested groups, the online algorithm gives a MAE on the order of 2 to 2.5 mph across all stations, which is applicable in developing a proactive traffic control system.

Prediction Confidence Interval Performance

The evaluation of confidence intervals is complicated by multiple objectives. In this study, two performance measures are used: (*a*) kickoff percentage, defined as the percentage of real observations falling outside the corresponding prediction confidence intervals, and (*b*) average confidence interval (CI) width-to-level ratio, defined as the average of CI widths divided by speed level. Ideally, the kickoff percentage will be close to the nominal significant level, for example, 5% for a 95% confidence interval, and the narrower the CI width-to-level ratio, the more informative the prediction confidence intervals. [See Kang and Schmeiser (*36*) and Schmeiser and Yeh (*37*) for more information.]

The prediction confidence interval performance of online algorithm and batch processing is presented in Figure 3. The performance for congested and uncongested categories is given separately.

Similar to forecasting accuracy performance, several observations can be drawn for the patterns of prediction confidence interval performance. First, for speed >30 mph, batch processing outperformed the online algorithm in both measures, that is, batch mode processing generates kickoff percentages closer to 5% than does the online algorithm, and batch mode generates lower prediction CI width-to-level ratio.

Second, for speed <30 mph, the online algorithm consistently outperformed batch mode across all stations in regard to CI width-to-level ratio, that is, the online algorithm gives narrower confidence intervals when traffic is congested. For kickoff percentage, the online algorithm slightly outperformed batch processing: the two approaches had comparable kickoff percentages at four stations, that is, 400329, 401195, 401209, and 400865. The online algorithm outperformed batch processing at five stations, 401052, 400445, 400976, 400838, and 400430, whereas batch processing outperformed the online algorithm only at Station 400443. This result indicates an improved prediction confidence interval performance of the online algorithm for congested traffic.

Finally, although not at all optimal, both approaches can generate workable prediction confidence intervals around the predictions at

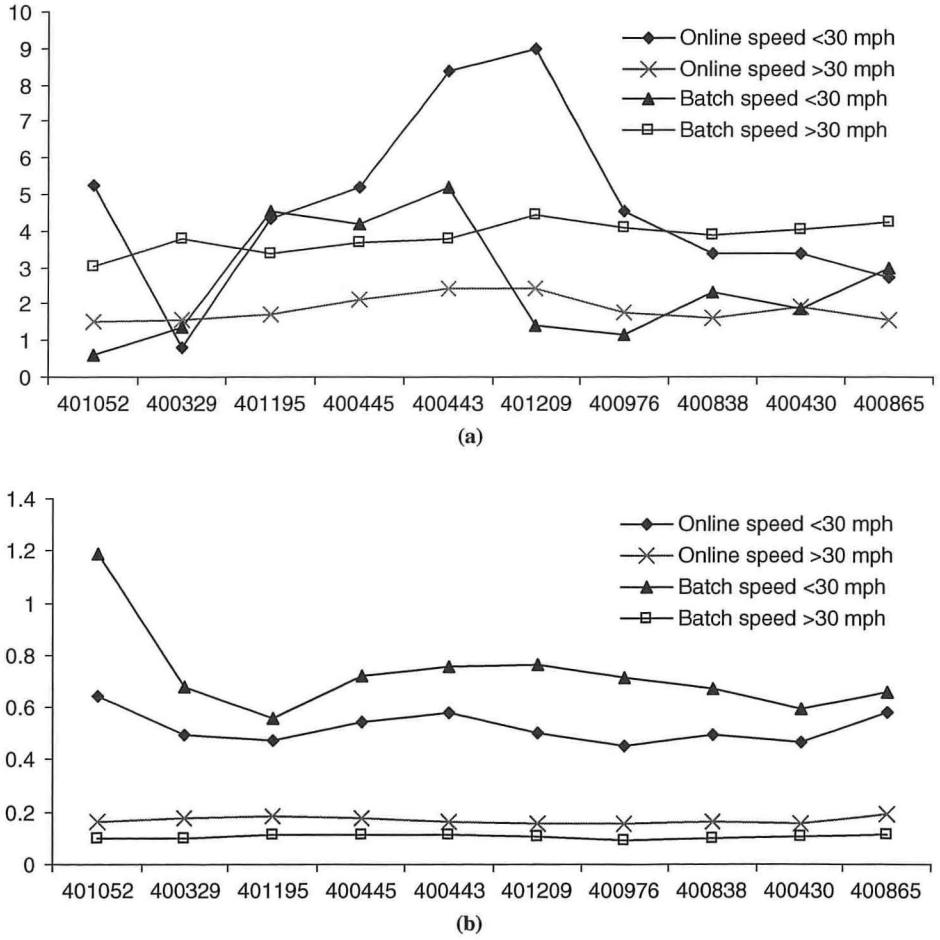


FIGURE 3 Prediction CI performance: (*a*) kickoff percentage and (*b*) CI width-to-level ratio (*x*-axis is station and *y*-axis is performance).

aggregated level, for example, the online algorithm produced a kickoff percentage closer to 5% at most stations.

Disaggregated Performance

The disaggregated performance of the online algorithm and batch mode processing is presented in Figure 4 for demonstrating the algorithm behaviors over a typical speed evolution pattern.

From Figure 4, first, the forecasts generated from the online algorithm and batch processing closely track the measured speeds for the congested and the uncongested traffic condition, indicating desirable speed level forecasting performance for both procedures.

Second, for speed >30 mph, batch processing generates narrower CIs than the online algorithm, which is consistent with the previous observation based on aggregated performance that batch processing outperformed the online algorithm for uncongested traffic in regard to prediction confidence intervals. Both batch processing and online processing generate CIs with reduced CI width variations for this group, indicating less volatile conditional speed variances for uncongested traffic.

Third, for speed <30 mph, the CIs from both procedures are widened when traffic is transitioning into and out of congestion, with a stronger widening effect for the batch processing than the online processing; in addition, the CI width from both procedures reduces with a higher reducing rate for online processing than for batch processing. On reflection, these different CI widening and reducing behaviors might demonstrate that the heteroscedasticity phenomenon is more pronounced for batch processing when the full series is used than for online processing when only data up to current time are used. In addition, the higher CI width reducing rate of online processing explains at least partially the observed behavior with aggregated performance that online processing outperformed batch processing for congested traffic in regard to prediction confidence intervals.

Finally, both procedures demonstrate workable prediction confidence intervals at the disaggregated level with measured speed falling out of the interval most likely occurring when traffic is transitioning into and out of congestion.

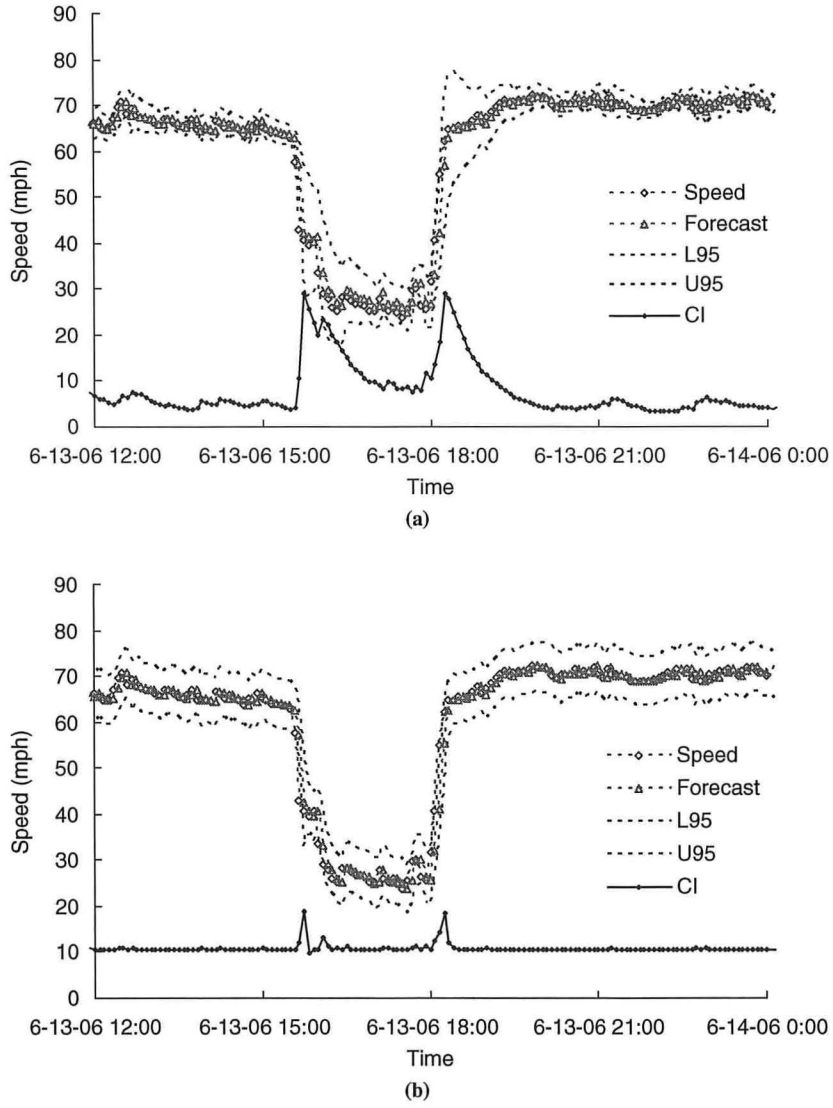


FIGURE 4 Disaggregated performance for afternoon peak hour traffic at Station 400329: (*a*) batch processing performance and (*b*) online processing performance.

CONCLUSION

It has been argued that short-term traffic condition forecasting, in regard to traffic level and traffic uncertainty, has the potential to contribute to alleviating congestion through proactive traffic control systems. Considering the desirable attribute of speed, that is, the one-on-one relationship between speed and roadway operating status, in this paper the traffic speed series is investigated by using a stochastic time series modeling structure, that is, the autoregressive moving average plus generalized autoregressive conditional heteroscedasticity (ARMA + GARCH) structure, in which the ARMA component is used to model the speed level evolution and the GARCH component is used to model the conditional speed variance evolution. In addition, the time series structure is processed into an online algorithm of layered Kalman filter structure using Kalman filters. Empirical investigation showed that workable performance, in forecasting accuracy and validity of prediction confidence interval, can be achieved through the proposed online algorithm. Because of the recursive nature of the Kalman filtering process, the computational demand of the proposed online algorithm is trivial with simplified field implementation.

Given the current situation on short-term traffic condition forecasting, future work could be envisioned as follows: (*a*) Development and test of a proactive traffic control scheme based on predicted traffic speed in a simulated environment. This can be considered as a step of validating and realizing the hypothesized benefits of a proactive traffic control system, which is long overdue in the literature of short-term traffic condition forecasting. Because this study is focused at the station level, the authors believe that a station-level proactive traffic control system, for example, ramp metering, could be selected for pilot testing. (*b*) Refinement of uncertainty modeling. Compared with traffic condition level modeling, conditional variance modeling is relatively less investigated in the literature. The ARCH model is promising in generating workable prediction confidence intervals as shown in this study and previous studies as well (*3, 17, 30*). However, extensive investigations are envisioned by the authors for improving and refining the conditional variance modeling approach. One study under way by the authors is to investigate the stochastic volatility model by incorporating the seasonality phenomena in the traffic condition series. In addition, uncertainty modeling and forecasting can be further extended to transportation link and network levels, thus facilitating the development of proactive traffic control systems for transportation corridors and networks.

ACKNOWLEDGMENT

This research was performed while the first author held a National Research Council Research Associateship Award at FHWA.

REFERENCES

1. Williams, B. M. *Modeling and Forecasting Vehicular Traffic Flow as a Seasonal Stochastic Time Series Process.* PhD dissertation. University of Virginia, Charlottesville, 1999.
2. Vlahogianni, E. I., J. C. Golias, and M. G. Karlaftis. Short-Term Traffic Forecasting: Overview of Objectives and Methods. *Transport Reviews,* Vol. 24, 2004, pp. 533–557.
3. Guo, J. *Adaptive Estimation and Prediction of Univariate Traffic Condition Series.* PhD dissertation. North Carolina State University, Raleigh, 2005.
4. May, A. D. *Traffic Flow Fundamentals.* Prentice Hall, Englewood Cliffs, N.J., 1990.
5. *Special Report 260: Strategic Highway Research—Saving Lives, Reducing Congestion, Improving Quality of Life.* TRB, National Research Council, Washington, D.C., 2001.
6. Williams, B. M., and L. A. Hoel. Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA: Theoretical Basis and Empirical Results. *Journal of Transportation Engineering,* Vol. 129, No. 6, 2003, pp. 664–672.
7. Ahmed, M. S., and A. R. Cook. Analysis of Freeway Traffic Time-Series Data by Using Box–Jenkins Techniques. In *Transportation Research Record 722,* TRB, National Research Council, Washington, D.C., 1979, pp. 1–9.
8. Hamed, M. M., H. R. Al-Masaeid, and Z. M. B. Said. Short-Term Prediction of Traffic Volume in Urban Arterials. *Journal of Transportation Engineering,* Vol. 12, No. 3, 1995, pp. 249–254.
9. Ross, P. Exponential Filtering of Traffic Data. In *Transportation Research Record 869,* TRB, National Research Council, Washington, D.C., 1982, pp. 43–49.
10. Nicholson, H., and C. D. Swann. The Prediction of Traffic Flow Volumes Based on Spectral Analysis. *Transportation Research,* Vol. 8, 1974, pp. 533–538.
11. Williams, B. M. Multivariate Vehicular Traffic Flow Prediction: Evaluation of ARIMAX Modeling. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1776,* TRB, National Research Council, Washington, D.C., 2001, pp. 194–200.
12. Kamarianakis, Y., and P. Prastacos. Forecasting Traffic Flow Conditions in an Urban Network: Comparison of Multivariate and Univariate Approaches. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1857,* Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 74–84.
13. Okutani, I., and Y. J. Stephanedes. Dynamic Prediction of Traffic Volume Through Kalman Filtering Theory. *Transportation Research Part B,* Vol. 18, No. 1, 1984, pp. 1–11.
14. Yang, F., Z. Yin, H. X. Liu, and B. Ran. Online Recursive Algorithm for Short-Term Traffic Prediction. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1879,* Transportation Research Board of the National Academies, Washington, D.C., 2001, pp. 1–8.
15. Shekhar, S. *Recursive Methods for Forecasting Short-Term Traffic Flow Using Seasonal ARIMA Time Series Model.* MS thesis. North Carolina State University, Raleigh, 2004.
16. Shekhar, S., and B. M. Williams. Adaptive Seasonal Time Series Models for Forecasting Short-Term Traffic Flow. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2024,* Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 116–125.
17. Guo, J., B. M. Williams, and B. L. Smith. Data Collection Time Intervals for Stochastic Short-Term Traffic Flow Forecasting. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2024,* Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 18–26.
18. Smith, B. L., and M. J. Demetsky. Short-Term Traffic Flow Prediction: Neural Network Approach. In *Transportation Research Record 1453,* TRB, National Research Council, Washington, D.C., 1994, pp. 98–104.
19. Chen, H., and S. Grant-Muller. Use of Sequential Learning for Short-Term Traffic Flow Forecasting. *Transportation Research Part C,* Vol. 9, No. 5, 2001, pp. 319–336.
20. Clark, S. D., H. Chen, and S. Grant-Muller. Artificial Neural Network and Statistical Modeling of Traffic Flows—The Best of Both Worlds. Presented at 8th World Congress on Transport Research, July 12–17, 1998, Antwerp, Belgium.
21. Dia, H. An Object-Oriented Neural Network Approach to Short-Term Traffic Forecasting. *European Journal of Operation Research,* Vol. 131, 2001, pp. 253–261.
22. Davis, G. A., and N. L. Nihan. Nonparametric Regression and Short-Term Freeway Traffic Forecasting. *ASCE Journal of Transportation Engineering,* Vol. 117, No. 2, 1991, pp. 178–188.
23. Smith, B. L., and M. J. Demetsky. Traffic Flow Forecasting: Comparison of Modeling Approaches. *ASCE Journal of Transportation Engineering,* Vol. 123, No. 4, 1997, pp. 261–266.
24. Smith, B. L., B. M. Williams, and R. K. Oswald. Comparison of Parametric and Non-Parametric Models for Traffic Flow Forecasting. *Transportation Research Part C,* Vol. 10, No. 4, 2002, pp. 303–321.

25. El Faouzi, N.-E. Combining Prediction Schemes in Short-Term Traffic Forecasting. *Proc., 14th International Symposium on Transportation and Traffic Theory* (A. Ceder, ed.), Jerusalem, Israel, 1999, pp. 471–487.
26. Van Der Voort, M., M. Dougherty, and S. Watson. Combining Kohonen Maps with ARIMA Time Series Models to Forecast Traffic Flow. *Transportation Research Part C,* Vol. 4, No. 5, 1996, pp. 307–318.
27. Guin, A. *An Incident Detection Algorithm Based on a Discrete State Propagation Model of Traffic Flow.* PhD dissertation. Georgia Institute of Technology, Atlanta, 2004.
28. Hugosson, M. B. Quantifying Uncertainties in a National Forecasting Model. *Transportation Research Part A,* Vol. 39, No. 6, 2005, pp. 531–547.
29. Sun, H., C. Zhang, B. Ran, and K. Choi. Prediction Interval for Traffic Time Series. Presented at 83rd Annual Meeting of the Transportation Research Board, Washington, D.C., 2004.
30. Kamarianakis, Y., A. Kanas, and P. Prastacos. Modeling Traffic Volatility Dynamics in an Urban Network. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1923,* Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 18–27.
31. Sohn, K., and D. Kim. Statistical Model for Forecasting Link Travel Time Variability. *ASCE Journal of Transportation Engineering,* Vol. 135, No. 7, 2009, pp. 440–453.
32. Tsekeris, T., and A. Stathopoulos. Real-Time Traffic Volatility Forecasting in Urban Arterial Networks. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1964,* Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 146–156.
33. Engle, R. F. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica,* Vol. 50, 1982, pp. 987–1008.
34. Bollerslev, T. Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics,* Vol. 31, 1986, pp. 307–327.
35. Kalman, R. E. A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME—Journal of Basic Engineering D,* Vol. 82, 1960, pp. 35–45.
36. Kang, K., and B. Schmeiser. Graphical Methods for Evaluating and Comparing Confidence-Interval Procedures. *Operations Research,* Vol. 38, No. 3, 1990, pp. 546–553.
37. Schmeiser, B., and Y. Yeh. On Choosing a Single Criterion for Confidence-Interval Procedures. *Proc., 2002 Winter Simulation Conference* (E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, eds.), 2002.

The Transportation Demand Forecasting Committee peer-reviewed this paper.

Supernetwork Approach for Multimodal and Multiactivity Travel Planning

Feixiong Liao, Theo Arentze, and Harry Timmermans

Multimodal and multiactivity travel planning is a practical but thorny problem in transportation research. This paper develops an improved supernetwork model to address this problem. The supernetwork is constructed mainly in three steps: a personalized network is first split into two types of networks with all links mode-specified; these are then assigned to all possible activity-vehicle states by means of state spreading from the beginning activity state. Finally, these discrete networks are connected into a supernetwork by state-labeled transition links. The proposed supernetwork is easier to construct than previous proposals and reduces the size needed to embody all combinations of choice facets explicitly. It can be proved for any activity program that any tour is a feasible solution in this representation. Consequently, every transport and transition link can be defined mode and activity-state dependent; thus standard shortest path algorithms can be used to find the most desirable tour. A case study is presented to show that the supernetwork model can be applied in a real-time manner for practical travel planning.

Multimodal trips are a common travel phenomenon and are expected to become more important because of their expected contribution to sustainable urban transportation. The multidimensional nature of multimodal trips makes it hard to model multimodal traveling (*1*). The complicated activity behavior in executing multiple activities renders travel planning more difficult. In this study, the focus is on the concept of a supernetwork to model transport networks and land use in an integrated fashion, and the model is used to analyze the extent to which it supports the planning and implementation of daily activity programs.

"Supernetwork" is defined as a network of transport networks integrating different transport modes (*2–4*). Links interconnecting the physical networks represent transfer locations where individuals can switch between modes. An example is a train station in which an individual can park a car and board a train. Such extended networks allow researchers to model multimodal trips as paths through the supernetwork. Nagurney and coworkers proposed further extensions of supernetworks that also include links representing particular transactions between actors and telecommunications to model transportation, communications, and transactions involved in supply chains and other economic activities (*5–7*).

Inspired by Nagurney and coworkers (*5–7*), Arentze and Timmermans (*8*) developed an extension of the basic supernetwork concept that integrates activity locations and multimodal transport networks. Their multistate supernetwork representation provides a potentially powerful framework for analyzing accessibility when accessibility is taken in its most fundamental meaning as the ease with which individuals can implement full activity programs. The cost of a least-cost path through a multistate supernetwork represents the effort associated with implementing an activity program. Such a measure takes into account multimodal and multiactivity patterns as well as the synchronization of transport networks and the land use system. A potential drawback of the approach is that the networks may become very large and possibly intractable because they need to incorporate as many copies of a physical network as there are possible states associated with the different stages of an activity program.

As Arentze and Timmermans argued, the approach may still be feasible when personalized supernetworks are constructed for one individual at a time (*8*). A personalized network allows not only representing preferences and perceptions individual specific but also a reduction to the relevant subset of a transport network. Thus, personalized supernetworks are not only more accurate, in the sense that they are tailored to the preferences and perceptions of an individual, but also reduce the size of the networks. This viewpoint has, however, not been validated, because a theoretical and quantitative analysis of supernetworks is lacking.

The purpose of this paper is to contribute to the further development of the supernetwork concept by providing such an analysis. Moreover, possibilities are explored of reducing supernetworks by improving the efficiency of the representation without reducing the representational power, and proofs of their proper working are provided. In doing so, the study makes a further step in the clarification of the theoretical properties and the operationalization of supernetworks for modeling and accessibility analysis of large-scale, integrated land use and transportation systems.

To achieve these objectives, the paper is structured as follows. First, basic concepts are briefly introduced and a formal description of a supernetwork model is presented. Next, the suggested improvements of the supernetwork representation are discussed and their properties are formally proved. A case study is carried out to indicate that the supernetwork model can be applied in a real-time manner for practical activity travel planning. Finally, a discussion of conclusions and future work ends the paper.

SUPERNETWORK MODEL

The supernetwork model is based on the fact that the costs of any kinds of links are mode and activity state dependent and personalized. State dependent means that link costs may vary with the current activity and mode state. Personalized refers to an individual's preference, perception, and knowledge of the links. In a supernetwork,

Urban Planning Group, Technische Universiteit Eindhoven, P.O. Box 513, 5600 MB Eindhoven, Netherlands. Corresponding author: H. Timmermans, H.J.P.Timmermans@tue.nl.

Transportation Research Record: Journal of the Transportation Research Board, No. 2175, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 38–46.
DOI: 10.3141/2175-05

the nodes denote real locations in space and every link represents an individual's action such as walking, cycling, driving, parking or picking up a car, boarding or alighting a bus or train, and conducting a specific activity. Thus, link costs can be readily defined state dependently and individually. The rest of this section discusses how a supernetwork based on these concepts is constructed. The following defines an activity program as a plan involving an individual leaving home with at most one private vehicle to conduct at least one activity and returning home with all activities conducted and all private vehicles at home.

Activity and Vehicle State

In this model of an activity program, every activity has only two states: either not conducted or conducted. An activity state is a possible combination of states across all activities. In practice, an activity might include several subactivities. For example, shopping may involve first shopping at a supermarket and then dropping the products somewhere. Such activities are decomposed into related activity units so that each of them involves a single location and a continuous time period. As a result, every activity has merely two states and there may be some implied sequence among the activity units belonging to a same broader activity. Therefore, if there are N activities, a possible activity state S^* can be described as N lengths of permutation of 0 and 1, $S^* = \ldots s_i^* \ldots, s_i \in \{0, 1\}$, where i is an index of activity and $s_i^* = 0$ denotes activity i not conducted or conducted.

Furthermore, the model allows different specifications of an activity program concerning flexibility of the activity sequence. For any two activities, if their sequence relationship is "immediate after," the sequence is "strict"; if just "after," it is "nonstrict"; otherwise, there is no sequence. If there is no sequence among N activities, the number of states $|S^*|$ equals 2^N. If there is a strict sequence among all activities whether inherently or individually, then $|S^*| = N + 1$. If the program includes two strict sequential parts, then $|S^*| \leq (N/2 + 1)$. In most real activity programs, N is a very small number. In most cases, N will not be larger than 3. Even if N may sometimes reach 5 or 6, the individual will consciously specify sequences on the basis of preference besides the inherent sequences (*9*). Thus, it is a safe assumption that in most situations the number of activity states is not larger than 20.

Simultaneously, a vehicle state defines where private vehicles are during the execution of the activity program. Because the individual goes out with at most one private vehicle, a possible state might fall into one of three situations: (*a*) all private vehicles stay at home, (*b*) the chosen private vehicle is in use, or (*c*) it is parked at a certain parking location outside. Therefore, a vehicle state $S^{\rightarrow}$ can be written as

$$S^{\rightarrow} = \ldots s_j^{\rightarrow}, \ldots, s_j^{\rightarrow} \in \{-1, 0, 1, 2, \ldots, p_j\}$$

where

j = index of private vehicle,
$s_j^{\rightarrow} = -1$ denotes that private vehicle j is staying at home,
$s_j^{\rightarrow} = 0$ private vehicle is in use, otherwise, parked somewhere, and
p_j = number of parking locations for j.

Hence, if there are M types of private vehicles and going out on foot is allowed, the number of possible vehicle states is given by

$$|S^{\rightarrow}| = 1 + \sum_j p_j + M$$

Assuming a three-way classification of going-out modes, an individual can go out on foot, by bike, or by using an available car. If by foot, no parking locations are involved; if by bike, the parking locations are normally designated to activity locations or transit locations near home; or else if by car, a robust heuristic is needed to reduce the choice set and find one or two parking locations near activity and transit locations. In general cases, for a chosen going-out mode, the number of vehicle states is within 2 times N.

The activity–vehicle state is the intersection of activity state and vehicle state, which demonstrates the situation in regard to which activities have been conducted and where the private vehicles are.

Multimodal Personalized Network

It is necessary to specify link costs state dependently, but it is redundant to consider the whole transport network. Given an activity program, only an activity-related subnetwork is useful for the individual, which is considerably reduced from the raw transport network. In Arentze and Timmermans, a single personalized network is extracted before the supernetwork is constructed (*8*). As shown below, the personalized network is further split, which can contribute to expressing the choice facets more clearly and reducing the scale.

Two types of networks are extracted in regard to going-out modes. One is the private vehicle network (PVN), which is accessible only by the chosen private vehicle. PVN contains the home location, parking locations, a few key locations, and links that connect all locations. Obviously, if the individual does not consider going out by private vehicle, a PVN is not needed. The other is the public transport network (PTN), which can be accessed by foot and other modes provided by public transport. PTN includes the home location, activity locations, parking locations, auxiliary transit locations, and mode-specified links that connect all locations.

PVN and PTN can be considered as bidirected and sparse graphs as they are extracted from road/service networks. Meanwhile they are connected because any nodes in PVN and PTN are reachable from home. Because PTN is a multimodal network, if any node induces a mode change, extra bidirected links are added to denote boarding–alighting transition links. For example (see Figure 1), Link 2→6 denotes boarding and Link 6→7 denotes alighting and then boarding. This extension seems to make the PTN large again. However, on the basis of the authors' observations, a PTN never has more than 40 nodes for three activities by pseudo-admissible heuristic extraction.

Extended in such a way, every link in PVN and PTN is mode specific. When copies of PVNs and PTNs are assigned to different activity states, PVNs and PTNs can be defined mode and activity state dependent.

Supernetwork Representation

To capture all choice facets for an activity program, the next step is to connect all PVNs and PTNs in different states through transition links, which cause entering different networks. A transition link represents parking–picking up a private vehicle or conducting an activity. Using the former implies an exchange between PVN and PTN, whereas using the latter leads to entering networks of different activity states.

If travel is not made by a private vehicle, no parking or picking-up transition link is involved. In the case of private vehicle m with

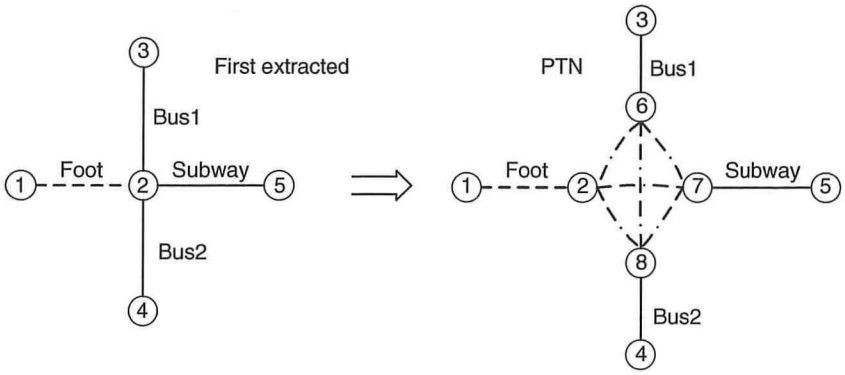


FIGURE 1 Example of extra links for mode change.

p_m parking locations, the transition between different vehicle states can be realized by one PVN and p_m PTNs. Links from PVN to PTN are parking links, and vice versa picking-up links (see Figure 2). The bold lines indicate that the individual picks up the private vehicle at parking location P_1, travels through PVN, and parks the private vehicle at parking location P_3 to conduct the next activity nearby. Compared with each parking or picking-up link resulting in a full reduced network in Arentze and Timmermans (*8*), a single PVN is added to erase all unnecessary copies of PTNs appearing in the vehicle state when a private vehicle is in use. Similarly, doing so reduces the size by erasing unnecessary copies of PVNs in the vehicle states when the private vehicle is parked.

Activity transition link occurs when any activity state alters from 0 to 1. Let S_k^* denote the set of activity states in which k activities have already been conducted, and $S_{k,m}$ $(1 \le m \le |S_k^*|)$ represent the mth element of the set. If $S_{k+1,n}$ is reachable from $S_{k,m}$ by conducting activity i, there are activity transition links between these two states. In particular, if activity i can be conducted at l_i different locations, l_i links are added in each pair of PTNs appearing in one vehicle state and two activity states (see Figure 3). A straightforward way that exhibits all of the activity transition in the whole activity state space is to start from $S_{0,1}$ and spread transition links to S_1^*, then from S_1^* to S_2^*, and so on until S_{N-1}^* to $S_{N,1}$.

Another improvement of the proposed supernetwork representation is that it is constructed separately in regard to the choice of going-out modes. In Arentze and Timmermans (*8*), all possibilities are contained in one scheme, in which one least-cost path will be found. In fact, links from different going-out mode induced networks can never be in one feasible path. Thus, constructing them separately does not affect optimality. The least-cost algorithm can be implemented in each going-out mode–based supernetwork, which can not only output different going-out mode specific least-cost paths, but also surprisingly cut down computing costs in real-time settings given the fact that there is no absolute linear-time shortest path algorithm so far.

On the basis of the components analyzed above, for each going-out mode, the steps for a supernetwork representation can be described as follows:

Step 1. Extract PVN and PTN, and extend PTN with boarding and alighting links, $k = 0$.

Step 2. For every activity state in S_k^*, connect PVN and PTNs by parking or picking-up links if PVN exists.

Step 3. For any activity state in S_{k+1}^*, if it can be reached by one from S_k^*, connect PTNs by activity transition links.

Step 4. $k = k + 1$, if $k < N$, go to Step 2; otherwise, stop.

Thus, the union of all the going-out mode-based supernetworks is the final supernetwork. Figures 4*a* and 4*b* are illustrations of an activity program that includes two activities and two going-out modes, that is, by foot and car. H and H′ denote home at the beginning and ending activity state, respectively; A_1 and A_2 denote locations for Activity 1 and 2; and P_1 and P_2, parking locations for the car. The bold tour in Figure 4*b* represents the tour in which the individual leaves home by car, parks car at P_1, and travels in PTN to conduct A_1; then picks up car at P_1, drives car again, parks at P_2, and travels in PTN to conduct A_2; and last picks up car at P_2 and returns home with all activities conducted. Along this tour, every link denotes a unique action and all choice facets are explicit.

It can be observed that all PTNs in the same activity state seem identical, whereas PTNs from different activity states tend to be different. However, merging the same PTNs into one brings the risk of contradictory tours. For example (see Figure 5), the tour marked with the bold links is infeasible; the individual cannot pick up the car at P_2 because it is parked at P_1. It is because of these different PTNs coupled with other components that a supernetwork can embody all choice facets concerning multimodal and multiactivity travel.

The supernetworks are constructed separately in regard to going-out modes. Therefore, all the going-out mode-based supernetworks possess the same characteristics. In each of them, it is argued that any path P from H to H′ is a feasible solution to the multimodal and multiactivity travel planning problem.

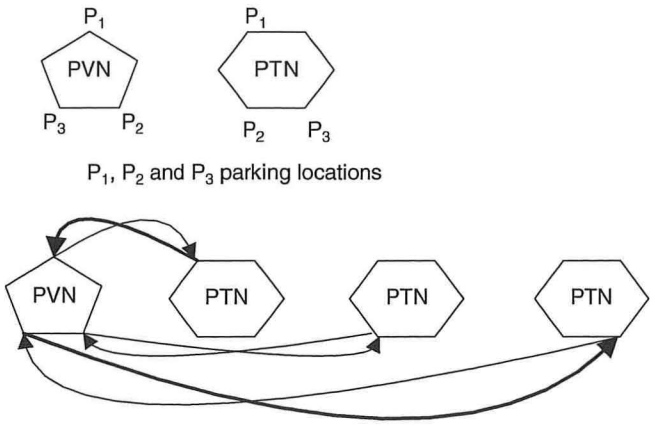


FIGURE 2 Example of parking or picking-up links.

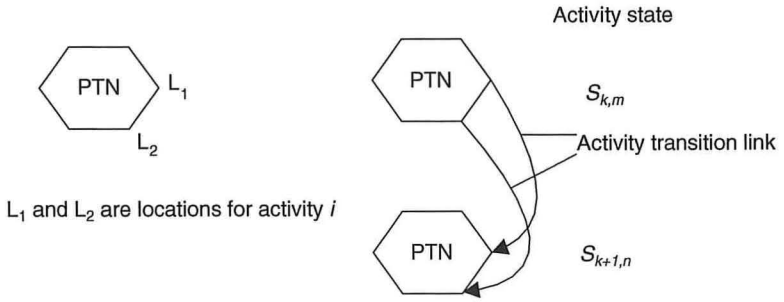
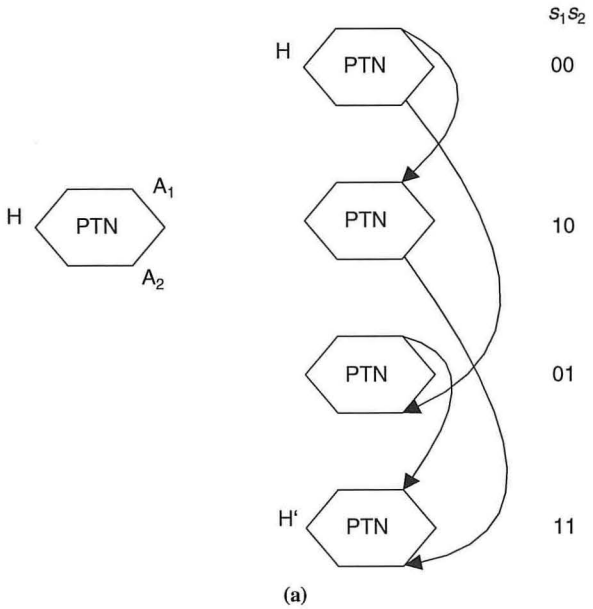
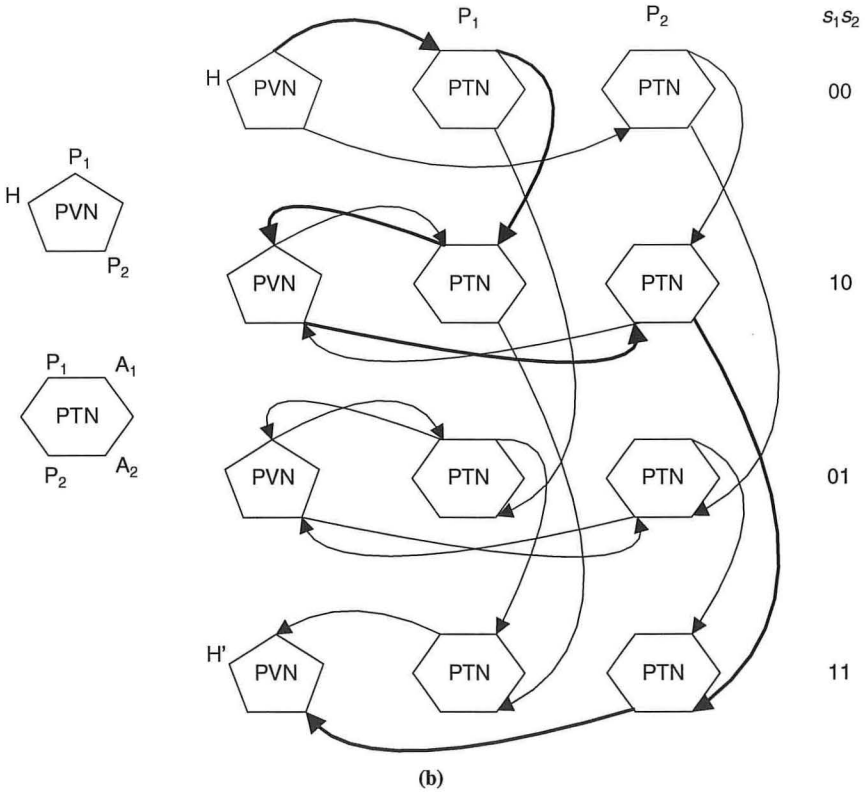


FIGURE 3 Example of activity transition links.



(a)



(b)

FIGURE 4 Supernetwork representations: (*a*) by foot and (*b*) by car.

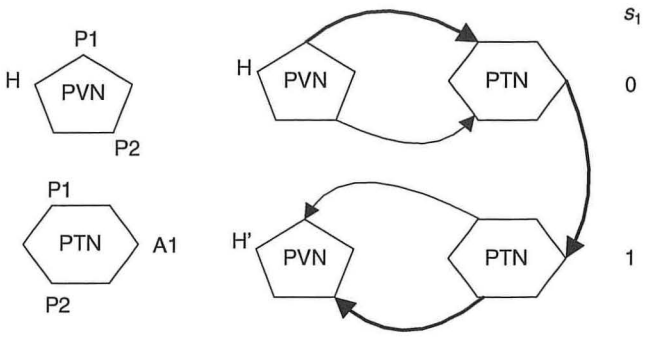


FIGURE 5 Example of infeasible tour.

Lemma 1

In every going-out mode-based supernetwork, there exists at least one path from H to H′; furthermore, any path P from H to H′ is feasible.

Proof. Consider the private vehicle mode first. In each activity state, PVN and PTNs are connected by parking or picking-up links at parking locations. Because PVN and all PTNs are connected, the horizontal units of a supernetwork are connected. Similarly, there are transition links between reachable PTNs. The last activity state must be reachable by the first one; otherwise, the activity program is erroneous. Therefore, the vertical units of a supernetwork are connected. In sum, the whole supernetwork is connected and thus there exists a path at least from H to H′.

A feasible path satisfies two conditions: (*a*) no contradiction in activity sequence relationships and parking–picking logic along the path and (*b*) all activities have been conducted and the private vehicle at home is at the end H′.

During the supernetwork construction, activity transition links occur only when activity states are one way reachable, so that the activity sequence relationship is naturally satisfied. In addition, in every activity state, PTNs are independent and correlated only by means of PVN. To conduct an activity, the individual must have the private vehicle parked in PVN first and enter a PTN specified by the corresponding vehicle–activity state. Once the activity is conducted, the activity state is updated. If it is the final activity state, the individual will pick up the private vehicle in PVN and return home. Otherwise, the individual has two options to conduct the next activity: either staying in the PTN of the same vehicle state or entering another PTN of a different vehicle state by going through the PVN. The whole process ensures that no conflict of parking or picking-up logical relationship will occur.

The endpoint H belongs only to the final activity state with all activities conducted, and it can be accessed only through the final PVN so that the private vehicle must be at home in H′. Therefore, any path from H to H′ is feasible.

Similarly, if by foot, there is no PVN and only one PTN, the argument still holds.

end of proof

Size of Supernetwork Representation

The nice properties of a supernetwork come at the cost of a substantial increase in the scale of networks. However, it is not difficult to calculate the size of the supernetwork for an activity program because all links and networks are well-ordered as activity × vehicle state matrices. Assume that the sizes of personalized networks are constant, then the size of a supernetwork depends on how many copies of the personalized networks and transition links there are.

Consider an activity program with N activities, l_i activity locations for activity i, M types of private vehicles plus by foot mode, and p_j parking locations for private vehicle j. If there is no sequence among activities, the number of copied networks Q_c is

$$Q_c = 2^N \times \left[1 + \sum_{j=1}^{M} (1 + p_j)\right] \tag{1}$$

where 2^N is the number of activity states and the rest is the number of vehicle states. This formula can be reduced to $|S^*| \times |S^{\rightarrow}|$. The number of parking or picking-up transition links Q_p is

$$Q_p = 2 \times (2^N - 1) \times \sum_{j=1}^{M} p_j \tag{2}$$

The reason for decreasing 1 is that there are only parking links in the first activity state and only picking-up links in the last. The number of activity transition links Q_a is

$$Q_a = \sum_{i=1}^{N} l_i \times 2^{N-1} \times \left(1 + \sum_{j=1}^{M} p_j\right) \tag{3}$$

These calculations are related directly to the sequences of activities. If specifying strict sequences for all activities by index, then

$$Q_c = (N+1) \times \left[1 + \sum_{j=1}^{M} (1 + p_j)\right] \tag{4}$$

$$Q_p = 2 \times N \times \sum_{j=1}^{M} p_j \tag{5}$$

$$Q_a = \sum_{i=1}^{N} l_i \times \left(1 + \sum_{j=1}^{M} p_j\right) \tag{6}$$

The formulas are not as simple as above when partly strict or nonstrict sequences are specified, but it is certain that they are somewhere in between these two situations. Taking the case in Arentze and Timmermans for example (*8*), N_1 and N_2, activities without and with product, respectively, there are $N_1 + 2 \times N_2$ activities after activity decomposition. If $l_i = 1$ for all i, the formulas are

$$Q_c = 2^{N_1} \times 3^{N_2} \times \left[1 + \sum_{j=1}^{M} (1 + p_j)\right] \tag{7}$$

$$Q_p = 2 \times \left(2^{N_1} \times 3^{N_2} - 1\right) \times \sum_{j=1}^{M} p_j \tag{8}$$

$$Q_a = \sum_{k=1}^{N_1+N_2} T(k) \times k \times \left(1 + \sum_{j=1}^{M} p_j\right) \tag{9}$$

where

$$T(k) = \sum_{\max(0,k-N_2)}^{\min(k,N_1)} \frac{N_1!}{i! \times (N_1 - i)} \times \frac{N_2!}{(k-i)! \times (N_2 - k + i)!} \times 2^{k-i} \tag{10}$$

The original problem for multimodal and multiactivity travel planning can be reduced to the traveling salesman problem in polynomial time, which is a famous NP-complete problem in combinatorial optimization. In other words, the original problem belongs to the NP-hard class. Fortunately, in reality, not every NP-hard problem is really that hard.

ACTIVITY TOUR FINDING

In the supernetwork, any node denotes a real location, and any link is either a transport link, which always causes a change of location, or a transition link, which never causes a change of location but a change of mode or activity state. Combined with the fact that links in PVN or PTN are all mode specific, each transport link has its own activity state and mode, and each transition link has its activity state rather than mode.

The generalized link cost pattern in Arentze and Timmermans is adopted (*8*) (which reveals the disutility on all links) for transport link, described simply as

$$\mathrm{cp}_l = f\left(\bar{\omega}_{m,s}(l), t_l, d_l, p_l\right) \tag{11}$$

where cp_l is generalized cost of transport link l and $f(\bar{\omega}_{m,s}(l), t_l, d_l, p_l)$ denotes function of activity state, mode, distance time elapse, and road preference, respectively. Likewise, the link costs for transition links are defined as

$$\mathrm{cs}_n = h\left(\pi_s(n), t_n, c_n, p_n\right) \tag{12}$$

where cs_n is generalized costs of transition link j and $h(\pi_s(n), t_n, c_n, p_n)$ denotes a function of activity state, service time, service cost, and location preference, respectively.

As the functions above suggest, all link costs are state dependent. For each transport link, if the activity and mode state are known, so are the other parameters of the link cost. It signifies that transport link costs are only state dependent. Transition link costs can also be recognized as only state dependent if other parameters are thought of as state dependent. This assumption is logical and possible as long as the individual specifies previous expected values to service costs and time. With all link costs only state dependent, the following can be argued.

Lemma 2
In each going-out mode-based supernetwork, if all link costs are only state dependent, the path P found by the Dijkstra algorithm is the least-cost path.

Proof. If link costs depend only on states, the costs of either transport or transition links are known in any known states. Because the supernetwork represents all feasible activity–vehicle states, all link costs in the supernetwork are known in advance. Given that link costs are defined as a disutility, link costs cannot be negative. Thus, the Dijkstra algorithm can find the least-cost path, and it is acyclic (*10*).

end of proof

Thus, the single-source (H) single-link (H′) shortest path algorithm fits the supernetwork model. Theoretically, the time complexity for the Dijkstra algorithm with binary heap is $O((m+n) \times \log n)$, where m and n denote the number of links and nodes; with Fibonacci heap, the time complexity is $O(m + n \times \log n)$ (*10*). Because PVN and PTN are both sparse, the supernetwork is also sparse with $m = O(n)$.

In addition, some service costs may also be time dependent because services are often distributed or associated with time. One special structural property concerning time-dependent links is called first-in, first-out (FIFO) (*11*). If all links in a network obey FIFO, the network exhibits the FIFO property, for which the label-setting method can also find the optimal tour. According to Lemma 2, if all links are only state dependent, the link costs are constant so that the supernetwork is a special case of an FIFO network. However, if any time-dependent link such as parking or boarding transition link brings the non-FIFO property, the supernetwork is a non-FIFO network, for which to find the least-cost tour is another kind of NP-hard problem. Fortunately, on the basis of some special reductions, a non-FIFO time-dependent link can be converted into FIFO again (*12*).

On the basis of the analysis of quantities of supernetwork components, an upper bound approach analysis case can explain the feasibility of the algorithm for practical use. Suppose that there are six activities, 20 activity states, 10 parking locations for one private vehicle, 20 nodes in PVN, and 80 nodes in PTN. Then, the number of nodes in one private vehicle-based supernetwork is 16,400 in total. For sparse graphs of such scale, the algorithm takes only a very small fraction of a second on a modern PC. Even with several choices of private vehicle, the whole computation time is within a second. In other words, the supernetwork model can react in a real-time manner for practical activity programs or can be applied in large-scale simulations.

All in all, the suggested supernetwork model suffices for general individual multimodal and multiactivity travel planning. Provided with a large set of real activity programs related to a simulated population, the supernetwork model can be tailored for accessibility analysis of integrated land use and transportation systems on a large scale for spatial or transportation planning.

CASE STUDY

In this section a case study is presented to indicate the efficiency of the supernetwork model for multimodal and multiactivity travel planning. The supernetwork model is executed in Matlab in a Windows environment running on a PC with Intel Core 2 Duo CPU E8400 @ 3.00 GHz 3.21G RAM. The case is selected from Arentze and Timmermans and concerns travel planning in the Almere–Amsterdam corridor of the Netherlands (*8*). Figure 6 is the personalized physical network, which is a symmetrical bidirected graph. For simplicity and without loss of generality, consider the case in which an activity program contains two activities (working, W, with one location and shopping, A, with two location alternatives) and one private vehicle (car with five parking locations, P), and that car is the only going-out mode considered and is the place for dropping off products.

Assume that the land use for activity locations and parking locations is as described in Table 1. Moreover, the disutility of boarding link at all stations is assigned a fixed quantity of five, and there is zero disutility for picking-up and alighting links, which are just marks of change of mode. Assume further that the activity state will not affect the disutility on the links, except that disutility will double on walking mode-specific links after shopping as a result of carrying bags, which is a reasonable assumption in daily life.

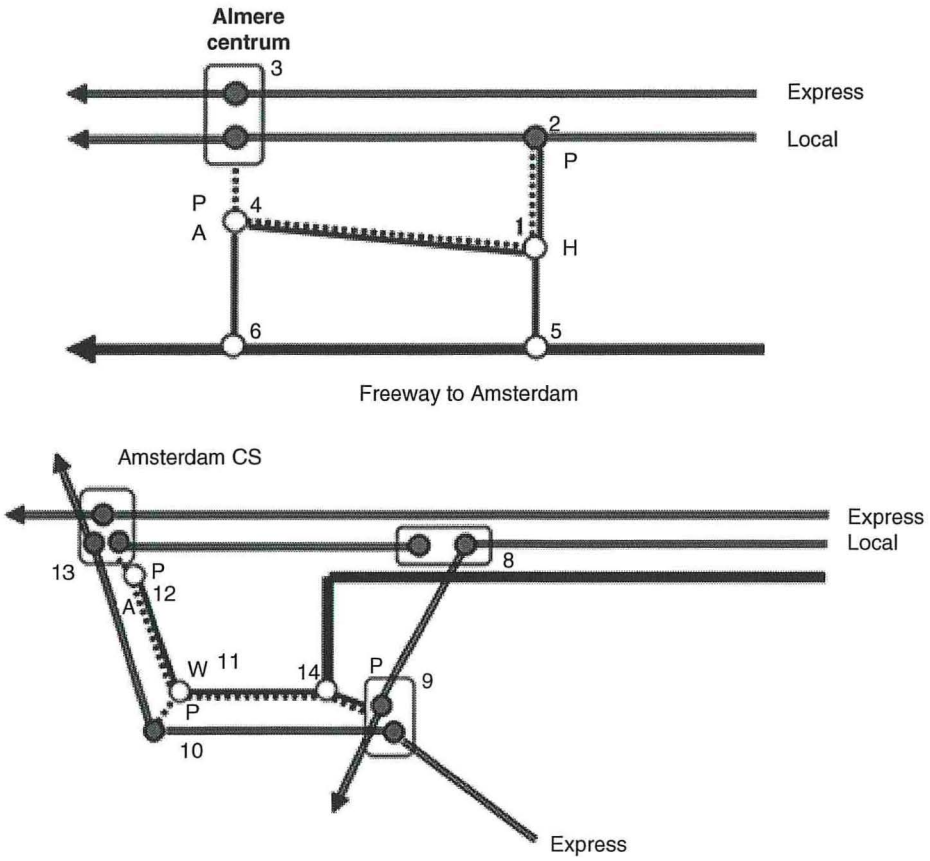


FIGURE 6 Almere–Amsterdam corridor.

According to the steps for constructing the supernetwork, PVN and PTN are first extracted from the personalized physical network. Figure 7*a* and 7*b* are PVN and PTN, respectively. The PTN is extended into hierarchical subnetworks marked by different modes, and boarding–alighting links are used to connect them. Let the number on each link denote the disutility at the first activity state. Because of space limitations, the remaining steps for connecting activity states by transition links are not shown.

After the supernetwork is constructed, the link costs (disutility) are to be assigned state dependently as assumed above. The running time for this activity program is 0.004 s. The optimal tour is listed in Table 2, which carries every detail of the activity–travel pattern.

In Table 2, the first two columns give the optimal tour for the activity program, and the last column gives the disutility on each link. The total disutility for the tour is 716. If the person buys only a few products and that does not affect the link cost on the walking links when the products are carried, what will happen? In that case, there is no need to reconstruct the supernetwork. After redefining the link cost and running the algorithm again, it is found that the optimal subtour

TABLE 1 Information on Land Use

Location	Service	Search Time	Cost	Preference	Time (min)	Disutility
1	Home	—	—	—	—	—
2	Parking	Short	Free	Low	2	10
4	Parking	Medium	Low	Low	4	24
9	Parking	Short	Free	Low	2	10
11	Parking	Medium	Free	High	4	16
12	Parking	Long	High	Medium	6	36
4	Shopping	Long	High	Low	45	135
12	Shopping	Short	Low	High	30	60
11	Working	—	—	—	9×60	540
Car	Dropping	—	—	—	0	0

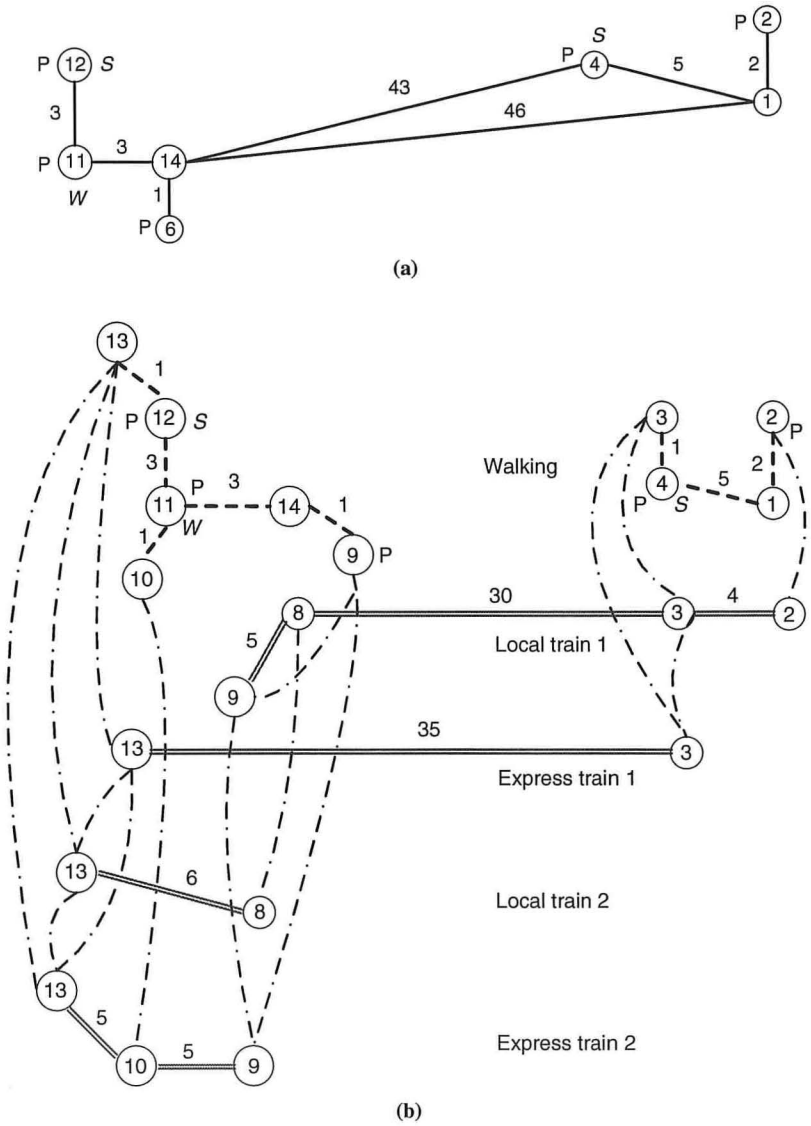


FIGURE 7 PVN and PTN: (*a*) extracted PVN for car and (*b*) extracted PTN with boarding and alighting links.

within the bold part has changed to another. In detail, after alighting at Station 3, the individual will not board Line 1 but will walk directly to the parking location (Node 2) through links (3, 4), (4, 1), and (1, 2). The total disutility on the new tour is 714. If the person changes the disutility again, the algorithm will again react in a real-time manner and provide the optimal tour.

CONCLUSION AND FUTURE WORK

In this paper the formal properties of the supernetwork model were analyzed, and the upper bounds of the size depending on assumed characteristics of the activity programs were derived. The analysis indicated that the size for personalized supernetworks stays well within reasonable bounds for realistic dimensions of activity programs. Furthermore, methods were developed to reduce the size of supernetwork representations without compromising the representational possibilities. It was shown that efficiency can be improved significantly so that larger problems can be handled with the same computing capacity. The approach based on a realistic case of a multimodal and multiactivity program was illustrated. Thus, the approach is applicable. The paper has made a next step in developing operational supernetwork models for accessibility analysis. Remaining steps concern the representation of time-dependent and time-window services that can represent the constraints of the transport and land use system and the definition of link cost functions that can represent actual preferences and rules for the selection of relevant nodes and links for tailored supernetwork representations. These steps will be considered in future research.

ACKNOWLEDGMENT

The study was supported by the Netherlands Organization for Scientific Research (De Nederlandse Organisatie voor Wetenschappelijk Onderzoek).

TABLE 2 Optimal Activity–Travel Tour

Link						
Start Point	End Point	Transport Link (yes?)	Mode	Transition Link (yes?)	Behavior	Disutility
1 - PVN	2 - PVN	Yes	Car		Departing	2
2 - PVN	2 - PTN			Yes	Parking	10
2 - PTN	2 - PTN			Yes	Boarding	5
2 - PTN	3 - PTN	Yes	Local train 1		Transferring	4
3 - PTN	8 - PTN	Yes	Local train 1		Transferring	30
8 - PTN	9 - PTN	Yes	Local train 1		Transferring	5
9 - PTN	9 - PTN			Yes	Alighting	0
9 - PTN	14 - PTN	Yes	On foot		Transferring	1
14 - PTN	11 - PTN	Yes	On foot		Transferring	3
11 - PTN	11 - PTN			Yes	Working	540
11 - PTN	12 - PTN	Yes	On foot		Transferring	3
12 - PTN	12 - PTN			Yes	Shopping	60
12 - PTN	13 - PTN	Yes	On foot		Transferring	2
13 - PTN	13 - PTN			Yes	Boarding	5
13 - PTN	3 - PTN	Yes	Express train 1		Transferring	35
3 - PTN	3 - PTN			Yes	Alighting and boarding	5
3 - PTN	2 - PTN	Yes	Local train 1		Transferring	4
2 - PTN	2 - PTN			Yes	Alighting	0
2 - PTN	2 - PTN			Yes	Dropping	0
2 - PTN	2 - PVN			Yes	Picking	0
2 - PTN	1 - PVN	Yes	Car		Returning	2

REFERENCES

1. Fiorenzo-Catalano, S. *Choice Set Generation in Multi-Modal Transportation Networks.* PhD dissertation. Delft University Press, Delft, Netherlands, 2007.
2. Sheffi, Y. *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods.* Prentice Hall, New Jersey, 1985.
3. van Nes, R. *Design of Multimodal Transport Networks: A Hierarchical Approach.* PhD dissertation. Delft University Press, Delft, Netherlands, 2002.
4. Carlier, K., S. Fiorenzo-Catalano, C. Lindveld, and P. Bovy. A Supernetwork Approach Towards Multimodal Travel Modeling. Presented at 82nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2002.
5. Nagurney, A., and F. Smith. Supernetworks: Paradoxes, Challenges and New Opportunities. *Proc., 1st International Conference on the Economic and Social Implications of Information Technology,* Washington, D.C., 2003, pp. 229–254.
6. Nagurney, A., J. Dong, and P. L. Moktharian. Teleshopping Versus Shopping: A Multi-Criteria Equilibrium Framework. *Mathematical and Computer Modeling,* Vol. 34, 2001, pp. 738–798.
7. Nagurney, A., J. Dong, and P. L. Moktharian. Multicriteria Network Equilibrium Modeling with Variable Weights for Decision Making in the Information Age with Application to Tele-Communicating and Teleshopping. *Journal of Economic Dynamics and Control,* Vol. 26, 2002, pp. 1629–1650.
8. Arentze, T. A., and H. J. P. Timmermans. A Multi-State Supernetwork Approach to Modeling Multi-Activity, Multi-Modal Trip Chains. *International Journal of Geographical Information Science,* Vol. 18, 2004, pp. 631–651.
9. Arentze, T., and H. J. P. Timmermans. Robust Approach to Modeling Choice of Locations in Daily Activity Sequences. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2003,* Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 59–63.
10. Ahuja, R., T. Magnanti, and J. Orlin. *Network Flows: Theory, Algorithms, and Applications.* Prentice Hall, New Jersey, 1993.
11. Dean, B. Algorithms for Minimum-Cost Paths in Time-Dependent Networks with Waiting Policies. *Networks,* Vol. 44, No. 1, 2004, pp. 41–46.
12. Luo, W., and P. Han. Study on Non-FIFO Arc in Time-Dependent Networks. *Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing-Copyright,* IEEE Computer Society Press, Vol. 2, Washington, D.C., 2007, pp. 305–310.

The Transportation Demand Forecasting Committee peer-reviewed this paper.

Comparison of Agent-Based Transit Assignment Procedure with Conventional Approaches

Toronto, Canada, Transit Network and Microsimulation Learning-Based Approach to Transit Assignment

Joshua Wang, Mohamed Wahba, and Eric J. Miller

The public transportation system, a key part of a multimodal transportation network, has been widely viewed as an efficient way to reduce road congestion and pollution. Public transportation planners use transit assignment models to forecast travel demand and service performance. As technologies evolve and smart transit systems become more prevalent, it is important that assignment models adapt to new policies, such as traveler information provision. This paper investigates three transit assignment tools that represent three approaches to modeling transit trip distribution over a network of fixed routes. These tools are the EMME/2 Transit Assignment Module (Module 5.35), commonly used by planners; Toronto, Canada, Transit Commission's transit assignment tool, MADITUC; and the newly developed Microsimulation Learning-based Approach to Transit Assignment (MILATRAS). These approaches range from aggregate, strategy-based frameworks to fully disaggregate microscopic platforms. MILATRAS presents a stochastic process approach (i.e., nonequilibrium based) for modeling within-day and day-to-day variations in the transit assignment process in which aggregate travel patterns can be extracted from individual choices. Although MILATRAS presents a different standpoint for analysis in comparison with equilibrium-based models, it still gives the steady state run loads. MILATRAS performs comparatively well with EMME/2 and MADITUC. In addition, MILATRAS presents a policy-sensitive platform for modeling the effects of smart transit system policies and technologies on passengers' travel behavior (i.e., trip choices) and transit service performance.

The urban transportation system is multimodal, allowing passengers the choice between personal vehicles, public transit, and other modes of travel. The importance of public transit is widely known for its economic and environmental benefits, as reflected by the large government initiatives in developing the transit network. Public transit is envisioned to provide an attractive alternative for mitigating traffic gridlock by implementing innovative service designs and deploying new smart systems and technologies for operations control and customer information. There has been a great deal of research done on transportation policies for highway infrastructure with intelligent transportation system (ITS) deployments, and similar research continues to develop for the effective analysis of transit ITS policies.

The existing methodologies for the evaluation of public transport policies and operations (transit assignment models) cannot represent realistic important features of the emerging public transport smart systems. Emerging information, communication, sensor technologies, and innovative transit operations control strategies are becoming critical elements of the viable, competitive public transit system. Advanced public transportation systems (APTS) and automated travelers information systems (ATIS), through a variety of data collection and communication capabilities, support improved operations planning and real-time transit operations management. Such information systems are designed to provide timely information to transit passengers on the conditions of the network, thus affecting travel choice behavior. Conventional transportation planning methods have serious limitations in evaluating the effects of information technologies because they are sensitive neither to the types of information that may be provided to travelers (i.e., lack of dynamic representation of the transport network) nor to the traveler's response to that information (i.e., lack of behavioral modeling that explicitly treats information provision). Bus rapid transit (BRT) and light rail transit (LRT) systems require the implementation of a variety of APTS applications; existing transit assignment models are therefore not adequate for representing BRT and LRT characteristics. Meanwhile, traffic assignment procedures have recently implemented a microsimulation approach to describe the detailed behavior of the transportation system. Learning-based algorithms, for modeling travel behavior with agent-based representation, have been shown to result in different and more realistic assignments. These advances present great opportunities for further advancing the state-of-the-art of transit assignment modeling.

Wahba and Shalaby presented the conceptual development of a modeling framework for a dynamic transit assignment procedure based on agent- and learning-based concepts, namely, MIcrosimulation, Learning-based Approach to TRansit ASsignment (MILATRAS) (*1*). MILATRAS implements a departure time and transit path choice

J. Wang, Department of Civil Engineering, University of Toronto, 35 St. George Street, Toronto, Ontario M5S 1A4, Canada. M. Wahba, Department of Civil Engineering, University of British Columbia, 6250 Applied Science Lane, Vancouver, British Columbia V6T 1Z4, Canada. E. J. Miller, Department of Civil Engineering, Cities Centre, University of Toronto, 45 Spadina Avenue, Suite 400, Toronto, Ontario M5S 2G8, Canada. Corresponding author: J. Wang, joshua.wang@utoronto.ca.

Transportation Research Record: Journal of the Transportation Research Board, No. 2175, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 47–56.
DOI: 10.3141/2175-06

model based on the Markovian decision process (*2*). [See Wahba for a detailed description of MILATRAS (*3*)]. When ITS deployment is considered, the proposed approach acknowledges the importance of maintaining explicit representation of information available to passengers as well as dynamic representation of service characteristics. It therefore allows for explicit modeling and evaluations of operational impacts of investing in new technologies (e.g., ATIS).

With the public transit network operated by the Toronto, Canada, Transit Commission (TTC) as a case study, this paper investigates the validation of MILATRAS in comparison with two other transit assignment models widely used in industry. For transit service planning, TTC uses a commercial transit assignment tool that implements the model for the analysis of disaggregate itineraries on a transit network (MADITUC) (*4*). For systemwide transportation planning, the city of Toronto and other regional planning agencies use the aggregate (zone-to-zone) transit assignment procedure provided by the EMME/2 software system (*5*). The objective of this investigation is twofold: to evaluate the goodness of MILATRAS' outputs with regard to real-world ridership information and assignment results from EMME/2 and MADITUC and to demonstrate the advantages of MILATRAS in regard to modeling predictive power, and sensitivity and flexibility to policy scenarios.

TRANSIT ASSIGNMENT APPROACHES

The transit assignment problem seeks to predict transit route loads and levels of services on a given transit network that consists of a fixed set of lines. Transit assignment procedures distribute a given travel demand on a network and attempt to model the interaction between the travel demand and the network supply. In this section the three transit assignment approaches used in this investigation are briefly described. An extensive literature review of transit assignment approaches is beyond the scope of this paper because of word limitations [see Wahba and Shalaby (*6*)].

Strategy-Based Approach: EMME/2

Mathematical formulations for the transit assignment problem were proposed in the late 1980s, rooted in the concept of the set of attractive lines by Chriqui and Robillard (*7*) and the treatment of the common lines problem by Le Clercq (*8*). Extending the work by Chriqui and Robillard (*7*), Spiess and Florian proposed a strategy-based approach to the transit assignment problem, developing a linear programming model and solution algorithm (*9*). For the strategy-based model, Spiess and Florian presented a mathematical model in which in-vehicle travel times increase as a function of passenger flows (*9*). This formulation is used in the popular EMME/2 software, which is one of the tools commonly used by urban transportation planners to conduct travel demand forecasting and systems analysis for road and transit networks. Although this model presents significant advancements in modeling transit demand, Spiess and Florian acknowledge that the model has important limitations (*9*). The main limitation is that waiting times at stops are not affected by transit volumes, which simplifies/ignores the effects of congestion.

A strategy is a set of rules guiding transit riders to their destinations by using information that becomes available while waiting. A strategy is a generalization of a path and accounts for the fact that passengers choose from a set of multiple paths, or a set of attractive lines, and not simply a direct path to their destinations. The set of attractive lines depends on the information provided to transit riders at each node. The only "information" provided to travelers is the route number of the next arriving transit vehicle. In EMME/2, transit riders are assumed to arrive randomly at stops and to board the next arriving transit vehicle from their set of attractive lines. Once the strategy is computed, the volume associated with each trip may be assigned on selected links and transit line segments. More detailed information on the strategy-based approach can be found elsewhere (*10*). The two versions implemented in EMME/2 are included in this study:

- Aggregate transit assignment. The standard transit assignment procedure (Module 5.35) in EMME/2 is an aggregate transit assignment. The model uses zonal demand to generate assignment outputs such as link flows and stop counts. In zonal assignment procedures, travelers in a particular zone are assumed to originate from the same point in the zone, known as the zone centroid. The zone centroid is connected with the rest of the transportation network through centroid connectors.
- Disaggregate transit assignment. One of the assumptions of aggregate transit assignment is that passengers in a zone all originate from the same location, which is unlike reality. To overcome that barrier, disaggregate transit assignment procedures allow for the modeling of individual trips in a transportation network. Although aggregate assignment procedures are less complex, properly modeled disaggregate assignment procedures have been shown to better capture travel behavior.

The strategy-based approach has been widely accepted as a way to model transit demand for transit service with high-frequency services, uncongested networks, punctuality of transit service, and no provision of passenger information. However, in reality, transportation networks have a wide range of services, from high frequency subway service to lower frequency bus services. Furthermore, the development of smart transit systems provides transit signal priority and real-time traveler information; such smart features cannot be adequately modeled/evaluated with the EMME/2 aggregate representation of transit service.

Fully Disaggregate Modeling Framework: MADITUC

MADITUC is a disaggregate transit assignment model with three major components:

1. Network definition module that provides detailed description of every route in the network,
2. Demand assignment module that assigns observed and simulated flows of transit users to a given network, and
3. Generalized analysis module, allowing the planner to perform a wide range of analyses of the data contained or generated in the modules above.

The assignment module contains a shortest path algorithm according to a well-calibrated impedance function. The generalized travel time is a weighted sum of in-vehicle time, waiting time, and transfer penalties. Waiting time is a function of line headways, regularity, and lower and upper bounds; transfer penalties are a function of modes and intermodal fares. Unlike other transit assignment models, MADITUC assigns observed trips to the actual routes taken,

which allows for detailed analysis of actual travel behavior. Additional model details generally are not publicly available, because MADITUC is copyrighted commercial software; however, the overall framework can be found elsewhere (*4, 11*). TTC uses MADITUC for all of its service design and planning exercises.

Similar to EMME/2, MADITUC implements an aggregate representation of the transit service. Because TTC is progressively deploying ITS technologies (such as transit signal priority) and planning to introduce new BRT–LRT services, the need for a new dynamic transit assignment framework with transit ITS modeling capabilities has been growing.

Microsimulation, Learning-Based Approach: MILATRAS

The modeling of service dynamics has been the focus of recent developments in the field of transit assignment modeling, a notable example being the schedule-based transit assignment approach by Nuzzolo et al. (*12*). The proper representation of service dynamics is important not only for capturing performance variability but also for evaluating transit ITS technologies, which requires detailed representation of dynamic service conditions. Because transit assignment deals with both supply and demand modeling, the emerging focus on dynamic service modeling requires a corresponding shift in transit demand modeling to appropriately represent the dynamic behavior of passengers and their responses to ITS technologies. For that reason, researchers have been motivated to explore more behavioral models that deal with the role of information, knowledge levels, and decision styles.

MILATRAS was developed for the modeling of day-to-day and within-day dynamics of the transit assignment problem. MILATRAS considers multiple dimensions of the transit path choice problem: departure time choice, stop choice, and route (or run) choice. MILATRAS represents passengers and their learning and planning activities explicitly. The learning process is concerned with the specification of different trip components (e.g., in-vehicle time, out-of-vehicle time, convenience measures). The planning process considers how experience and information about those components on previous days influence the choice on the current day. The underlying hypothesis is that individual passengers are expected to adjust their behavior (i.e., trip choices) according to their experience with the transit system performance as stored in a "mental model." Individual passengers base their daily travel decisions on the accumulated experience gathered from repetitively traveling through the transit network on consecutive days. Individual behavior, therefore, is modeled as a dynamic process of repetitively making decisions and updating perceptions according to a learning process. By repeatedly making a decision, individuals acquire knowledge (i.e., learn) about their environment and thereby form expectations about attributes of the environment. Individuals may make different choices over time and thus learn which of these choices is more effective in achieving particular goals.

TTC SYSTEM AND APPLICATION DATA

TTC provides service coverage within the city of Toronto, Ontario, which is considered the economic heart of the Greater Toronto Area (GTA). TTC's market share of all daily transit trips made in the GTA is about 78% (*13*).

In 2001 TTC operated 294 one-directional routes (or 147 two-directional routes) during the morning peak period (*TTC Report.* Service summary report. Service Planning Department of the Toronto Transit Commission, 2001, unpublished work). Some of these routes run on different branches that covered different segments, bringing the total of individual branches modeled to 480 branches. These branches served just about 10,000 stops during the morning period. The frequency of service over these branches spanned from high frequency service (2-min headway) to low frequency service (60-min headway), with different values representing medium frequency services. The TTC system operated four different types of service: traditional bus service, express bus service, LRT and streetcar service, and rapid rail transit and subway service.

The system boundary for this study is the TTC network in 2001. In particular, this study looks at all transit trips made in the morning peak period (6:00 to 9:00 a.m.) that use the TTC service. Wahba and Shalaby report on a full-scale implementation of MILATRAS by using the TTC system as a case study (*14*).

Demand Data: Transportation Tomorrow Survey 2001

The Transportation Tomorrow Survey (TTS) is a comprehensive travel survey that collects detailed demographic and travel information on all household members for 5% of the household population in the GTA and surrounding regions. Collected data on transit trips include trip start time, trip purpose, origin and destination geolocations, and the sequence of transit routes used in each transit trip, among others. The TTS was first conducted in 1986 and has been carried out every 5 years since that time. The data used in this application were obtained from the 2001 TTS database.

For this study, a transit origin–destination (O-D) matrix was extracted from the TTS 2001 records for trips that use the TTC service and start between 6:00 a.m. and 9:00 a.m. The total number of trips in the TTC O-D matrix is approximately 320,000, considering three modes of access and egress, namely, walk, auto-passenger, and auto-driver modes. Disaggregate data were obtained for individual choices, including start time of trip, station and transfer point, and route (or sequence of routes). The total number of disaggregate records (actual records surveyed by TTS) is 19,650, reflecting approximately 5% of the total number of trips in which totals are extrapolated by using zonal expansion factors.

For the standard aggregate transit assignment implementation in EMME/2, the expanded totals (transit O-D matrix) are input to the transit assignment module. For the EMME/2 disaggregate transit assignment procedure, individual trip data were used. MADITUC works by reassigning the observed individual trips to the actual routes taken; therefore, disaggregate data are used in the MADITUC assignment. The input to MILATRAS is the traditional O-D transit matrix. This O-D matrix is then converted into individual trip lists, each representing a passenger–agent; this final list is consistent with the original transit O-D matrix such that the aggregation over origin zones, destination zones, or both will result in the original O-D matrix [see Wahba and Shalaby for details (*14*)].

Network Data: Toronto Transit Commission

TTC is the largest transit service provider by ridership in the GTA. With varying supply and demand characteristics throughout the

network, TTC serves as a good system to validate transit assignment models. The TTC 2001 network has been coded in EMME/2, MADITUC, and MILATRAS.

The strategy-based approach uses the transit network representation in EMME/2. The transit service is modeled as a strongly connected network of a set of nodes, transit lines, and walk links. A transit line is generally defined as a sequence of nodes at which passengers may board and alight. Each line has a header section and a route itinerary section. The header section defines attributes that apply to the entire line (e.g., line name, headway). A route itinerary section is defined by a sequence of segments; each segment is defined by a "from node" and "to node" and several attributes (e.g., dwell time, transit time function, layover time).

In MADITUC, the transit network definition has a physical geometry (spatially referenced) and a level of service (fare, travel time, accessibility) specification. More precisely, the NETWORK is defined as a set of NODES and LINES, which are characterized by x-y coordinates, descriptors, and attributes (e.g., length, route, speed, mode, vehicle type).

In both approaches, transit services are represented by the frequency of each transit line. Therefore, traveler's wait time is assumed to be dependent on the headway of "attractive lines." This is an acceptable approximation of the average wait time only if headways are small and the schedule is reliable. When headways are large, informed travelers are likely to coordinate their arrival time to be close to the time the bus arrives. This can be misrepresented in both approaches by inaccurate total travel times for low frequency lines, resulting in potential bias in the choice of transit lines on parallel routes (as in the strategy-based approach) or bus stops (as in the MADITUC approach). Because of the nonexistence of explicit representation of the runs for each transit line, timetable coordination is simply ignored. Timetable coordination is a major characteristic of transit networks. Moreover, line capacity is not accounted for.

In MILATRAS, a mesoscopic model was developed to represent the dynamics of the transit service at the network level with a detailed representation of branch- and vehicle-level operations. The developed mesoscopic model represents the movement of each transit vehicle between stops as a function of the link speed, without the explicit representation of the general traffic. Meanwhile, it microscopically represents individual passenger alighting and boarding activities at each stop, including the interactions between passenger agents and between passenger agents and the transit network. The supply model acknowledges loading priorities at stops and represents congestion through fail-to-board handling. This is modeled as a "discrete-time, event-driven" simulation model, in which the simulation model clock advances every time step (e.g., second) and handles events as they occur at varying increments [see Wahba for a description of the network simulation model (*3*)]. This mesoscopic model allows the proper modeling of the dynamics of each route, with regard to run-by-run representation and passenger–network interactions, while accounting for network-level (evolving) effects.

ANALYSIS SCENARIOS

Two sets of data are used for model validation. The observed counts from the TTS 2001 Validation Report are used to compare model outputs with actual route counts for the morning peak period (*15*). Boardings and alightings are also examined for TTC subway stops.

EMME/2 Scenarios

Eight different scenarios were run in EMME/2. Scenarios were modeled at the zonal aggregation level (aggregate assignment, abbreviated "Agg") and on an individual trip level (disaggregate assignment, abbreviated "Dis"). Since the EMME/2 network includes the entire transit network in the GTA and the study area is for the TTC service only, assignments were also conducted for the full GTA transit demand ("Full"), in addition to the TTC-only demand ("TTC") used for MADITUC and MILATRAS runs. Further, two versions of the EMME/2 implementation were tested: one in which the time value of fares is included in the transit path impedance calculations [referred to as the "Version 3" (V3) model] and one that does not incorporate transit fare in the path impedance function [the so-called "Version 2" (V2) approach]. The eight EMME/2 scenarios are thus denoted as V2_Agg_TTC, V2_Agg_Full, V2_Dis_TTC, V2_Dis_Full, V3_Agg_TTC, V3_Agg_Full, V3_Dis_TTC, and V3_Dis_Full.

MADITUC Approach

Two different outputs were provided by the TTC Planning Department from MADITUC: observed and simulated runs. Observed runs are the results of calibrating the model to observed data, and simulated runs are simulations in MADITUC using the calibrated values.

MILATRAS Approach

Eight scenarios were considered for MILATRAS (abbreviated "MIL" in figures for brevity). The runs are divided into two main categories: 5% input and 100% input. The 5% expanded runs use 5% of the population to generate the path choices (abbreviated as "5EXP"). In these scenarios 5% of the population is modeled explicitly with 100% total expanded demand, and each passenger of the 5% represents a certain number of passengers, depending on the passenger's associated expansion factor. The 5% runs in MILATRAS are runs that use only the 5% data, without any expansion factors (abbreviated as "5"). One hundred percent runs use the full 100% population at the outset, and these data are assigned to a transit network (abbreviated as "100"). Each case considered the effect vehicle capacity constraints had on the study area. Scenarios that included vehicle capacity constraints and did not include these constraints were abbreviated as "VC" and "NOVC," respectively. Also, exact geocoded locations of trip O-D for the 100% demand were not available and were randomly generated by using land use maps based on trip purpose [see Wahba and Shalaby (*14*)]. These randomly generated locations, which still correspond to the TTS zonal demand, were also considered in the analysis. The random geocoded and exact TTS geocoded (for the 5% and 5% EXP demand) scenarios are abbreviated with "RND" and "TTS," respectively. The MILATRAS scenarios are 100_TTS_VC, 100_TTS_NOVC, 100_RND_VC, 100_RND_NOVC, 5EXP_TTS_VC, 5EXP_TTS_NOVC, 5_TTS_VC, and 5_TTS_NOVC.

RESULTS AND DISCUSSION

The results of this study produced TTC route loads and TTC subway stop loads. Route loads were compared with actual counts, and stop loads were compared with TTS data for subway station board-

ings and alightings. Percentages, instead of actual loads, were used in calculations because different models produced different totals. Another reason is that demand data are extracted for trips starting during the a.m. peak period, whereas actual counts reflect the number of boardings during the a.m. peak period. To evaluate the accuracy of route and stop load distributions, the following three values were used: square of the Pearson product moment correlation (RSQ), global relative error (GRE), and point mean relative error (PMRE). These values are defined as follows:

$$\mathrm{RSQ}^2 = \left(\frac{\sum (x - x^*)(y - y^*)}{\sqrt{\sum (x - x^*)^2 \sum (y - y^*)^2}} \right)^2$$

where

x^* = mean of simulated output,
x = simulated output,
y^* = mean of observed output, and
y = observed output.

$$\mathrm{GRE} = \frac{\sum_{i=1}^{n} |\text{observed} - \text{simulated}|}{\sum_{i=1}^{n} \text{observed}}$$

$$\mathrm{PMRE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(\frac{\text{observed} - \text{simulated}}{\text{observed}} \right)^2}$$

For overall model comparison, five simulations were considered. For EMME/2, the updated Version 3.0 network was used with the full GTA transit demand. The EMME/2 network includes transit network representation for the entire GTA and would therefore produce more realistic TTC loads. Version 3.0 contains the modifications including transit fare in the general disutility function. Aggregate and disaggregate simulations are included (i.e., EMME/2_V3_Agg_Full and EMME/2_V3_Dis_Full). There is only one simulation of MADITUC, which is also included for analysis (MADITUC_SIM). For MILATRAS, the 100% and 5% expanded simulations were used (MIL_100_TTS_VC and MIL_5EXP_TTS_VC). Vehicle capacity constraints were included to simulate reality as closely as possible. The random geocode simulation for the 100% was also included to illustrate the impact of exact geocode data (MIL_100_RND_VC).

MILATRAS was run on a PC (Windows XP Professional, 2.39 GHz, 3.25 GB RAM). The 100% sample run in MILATRAS takes about 4 min per iteration and converges in 40 iterations. The 5% sample run in MILATRAS takes 1.5 min per iteration and converges in 15 iterations. The aggregate and disaggregate transit assignments in EMME/2 take about 2 min and 17 min, respectively; the simulations were run on a Sun Workstation (Sun Fire V210, 1.34 GHz, 4.5 GB RAM). The MADITUC assignment takes just over 13.5 min and was run on a PC (Windows XP Professional, 2.99 GHz, 2.05 GB RAM).

TTC Route Loads

Figures 1 through 4 show overall model performance statistics. Figure 1 considers all submodes together (bus, streetcar, subway). In this case, all model outputs closely replicate the actual route loads, with the RSQ for all models being approximately equal to 1. GRE values in Figure 1 show that the EMME/2 aggregate assignment does not perform as well as the other disaggregate simulations. This illustrates the improved accuracy achievable in using disaggregate assignment procedures. There is also a distinct difference in PMRE between the TTS and RND simulations of MILATRAS. Although there is not that much difference in GRE, randomly assigning geocodes causes an increase in PMRE. This reflects the importance of exact geocodes for accurate prediction of individual routes. The MILATRAS 5% expanded simulation reflects a slightly higher GRE than the 100% simulation. The reason is that MIL_5EXP_TTS_VC models 5% of the population uniquely and assumes that 100% of the population makes the same choices as the 5%.

Looking at each submode provides interesting insight into each model. Figure 2 presents the model statistics for the subway submode. The RSQ values for the subway submode are relatively the same. MADITUC performed exceptionally well in comparison with the other models. This level of accuracy is attributed to MADITUC

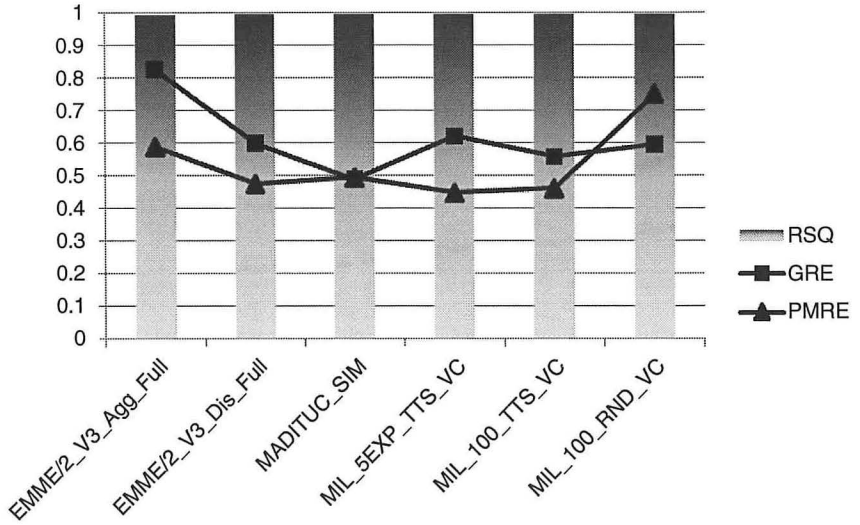


FIGURE 1 RSQ, GRE, and PMRE of TTC route loads: all modes.

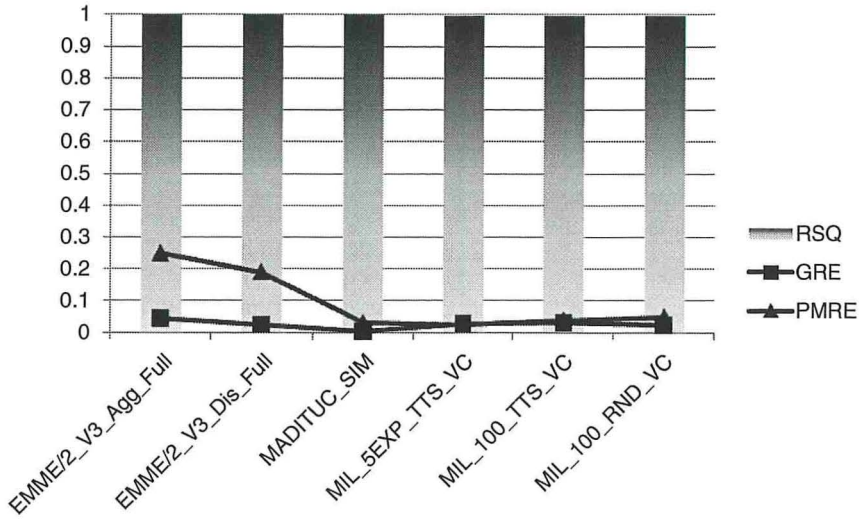


FIGURE 2 RSQ, GRE, and PMRE of TTC route loads: subway.

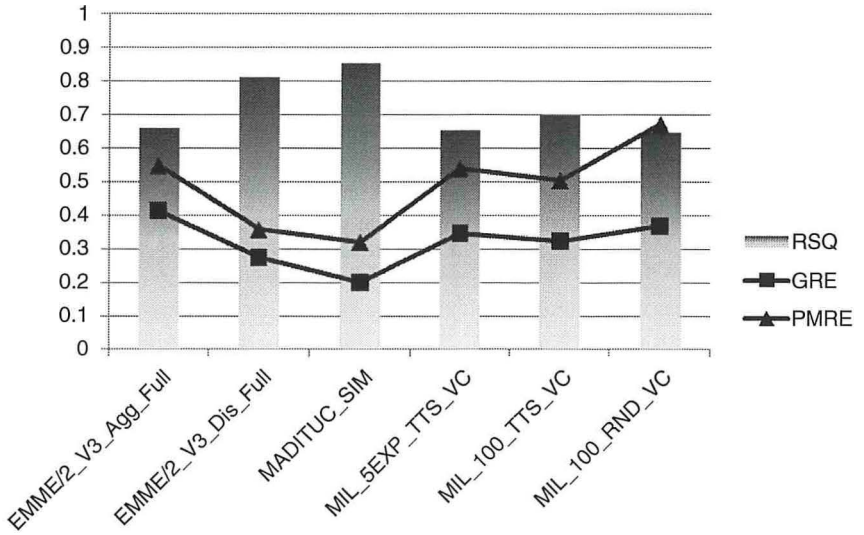


FIGURE 3 RSQ, GRE, and PMRE of TTC route loads: streetcar.

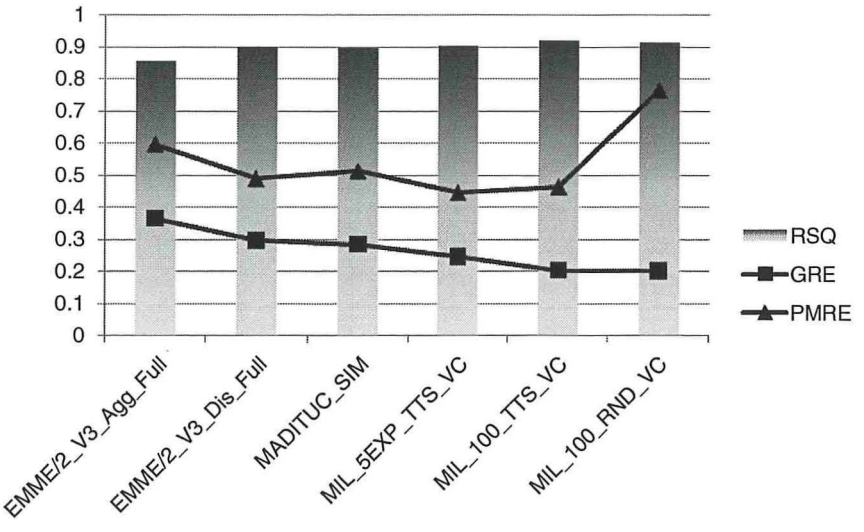


FIGURE 4 RSQ, GRE, and PMRE of TTC route loads: bus.

assigning observed trips to the actual routes taken, allowing for an accurate representation of service characteristics. There is also a marked difference in PMRE between EMME/2 and MILATRAS. MILATRAS uses microsimulation and agent-based modeling, representing the passengers and their learning and planning activities explicitly. Therefore, MILATRAS accounts for individual passenger choices and captures the subway demand better than the strategy approach implemented in EMME/2.

Figure 3 shows the statistics for the streetcar submode. In this submode, the RSQ varies more between models in this case. Figure 3 indicates that MADITUC outperforms the other models. In this case the disaggregate simulations using exact geocodes (MADITUC and MILATRAS) capture the data better than the EMME/2 standard transit assignment (aggregate). For MILATRAS, random geocode produces higher PMRE than exact geocodes. This reflects the importance of having exact geocodes for analyzing individual routes. The three MILATRAS simulations perform relatively the same on a systemwide scale, as reflected by the common RSQ and GRE values. There is no marked difference between the 5% expanded and 100% runs in MILATRAS at the streetcar submode level.

Figure 4 indicates that all models closely reproduce the actual bus demand. The disaggregate assignments produce better outputs than the EMME/2 aggregate assignment. An exception to this is the random geocode simulation in MILATRAS, which had a higher PMRE than all other models. The advantage of exact geocodes for MILATRAS is seen at an individual level (PMRE); however, at the systemwide scale, the choice of random geocodes does not affect the model performance (as reflected in the RSQ and GRE values). In addition, MILATRAS uses a detailed representation of all bus route stops (about 10,000 bus stops); with the random geocodes, stop choice may change, which affects the route choice.

In general, streetcar route loads experience more variability than other modes. Figure 5 shows the percent differences with respect to observed counts and associated headways for the TTC streetcars for EMME/2 disaggregate assignment (EMME/2_V3_Dis_Full) and the MILATRAS 100% simulation (MIL_100_TTS_VC). In this figure both models overestimate the 501 Queen streetcar and underestimate the 504 King streetcar. However, these two streetcar routes run in parallel lines, approximately 300 m apart, and have approximately the same combined demand as the actual counts. The 501 Queen has a mean speed of 10.3 mph and headway of 5.1 min. The 504 King has a mean speed of 8.8 mph and headway of 4.0 min; 501 Queen and 504 King have relatively the same headway. One explanation for the overestimation in the 501 Queen is that the Queen streetcar provides a higher level of service, traveling 1.5 mph faster than the King streetcar. Because these are parallel lines, the model may assign more passengers on the Queen streetcar instead of the King streetcar. In MILATRAS, passengers are picking the 501 Queen more in the mental model (i.e., choice set generation process). King Street also has a lot of businesses and hence attracts more trips in the a.m. peak period. This information is not included in the land use maps and hence not captured in the path generation model in MILATRAS. EMME/2 does not include land use characteristics in the path choice model, and as a result, the models underestimate the King streetcar. In addition, an underestimation in King and St. Andrew subway station alightings would also lead to an underestimation of the King streetcar boardings. The King station and St. Andrew station are both on King Street, located in the central business district and with direct connections to the King streetcar.

TTC Subway Stop Boardings and Alightings

In addition to route load outputs, TTC subway stop loads were considered to examine how models predicted boardings and alightings at individual stops, because different boarding and alighting combinations can yield identical route load outputs. The results show that MADITUC outperforms all other models because it assigns observed TTS trips to actual routes taken. Figures 6*a* and 6*b* show percentage differences for TTC subway stop boardings and alightings, respectively, for the EMME/2_V3_Dis_Full and the MIL_100_TTS_VC results. The figures show that MILATRAS predicts well subway stop boardings and alightings, especially in the downtown area where some stations are operating at critical capacity levels.

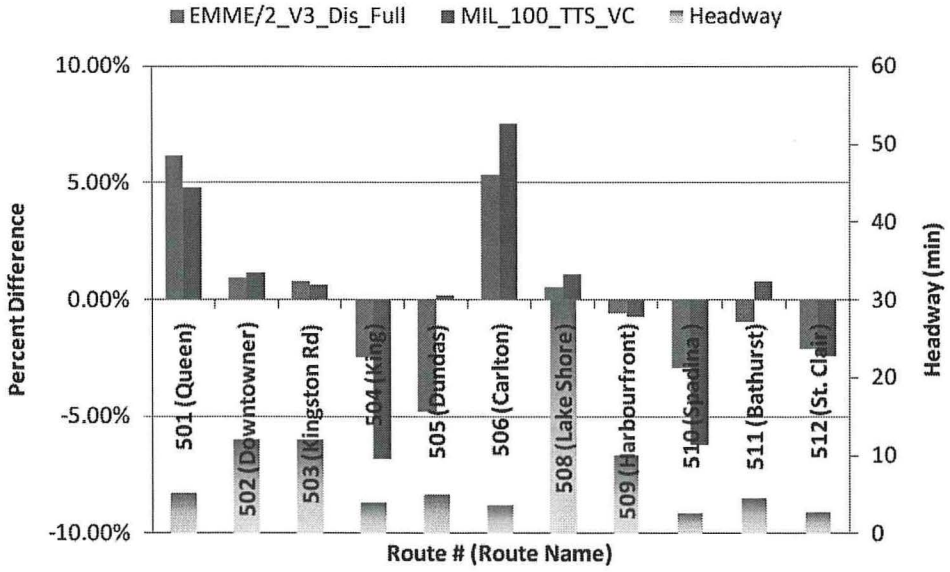
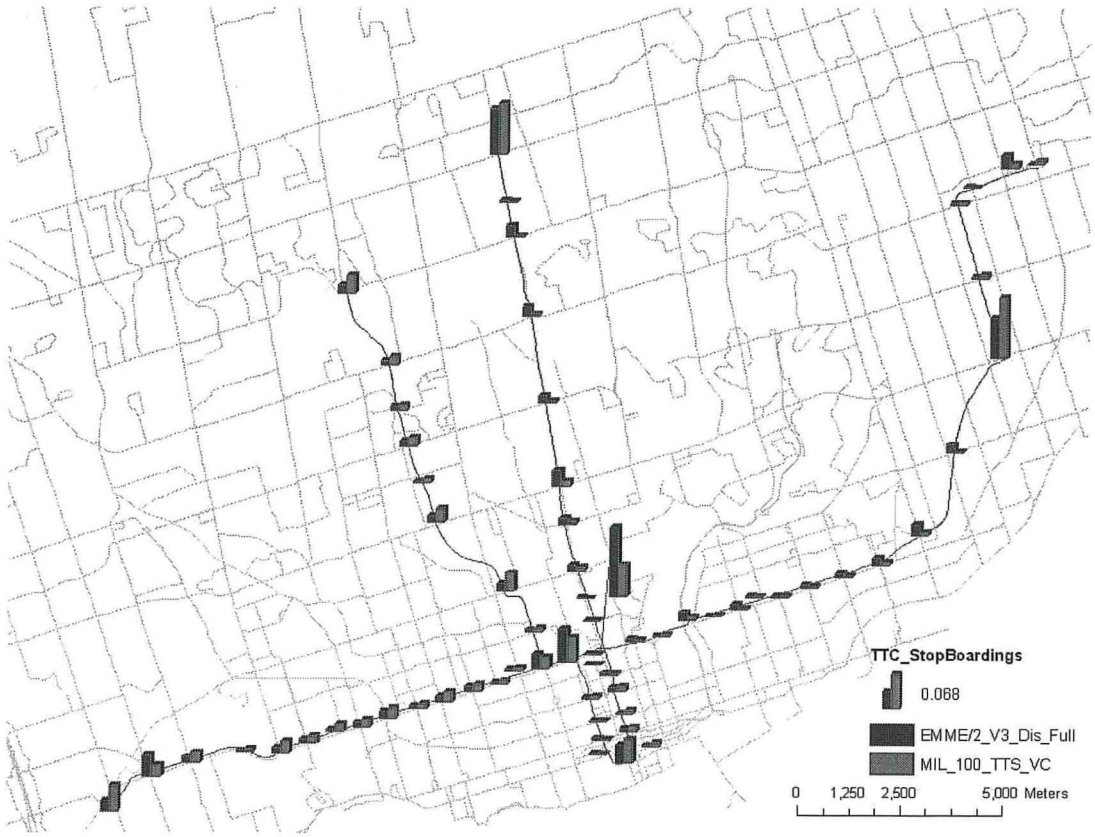


FIGURE 5 TTC streetcar route load comparison between EMME/2_V3_Dis_Full and MIL_100_TTS_VC.



(a)

(b)

FIGURE 6 TTC subway stop loads: (*a*) boardings and (*b*) alightings.

The results of the EMME/2_V3_Dis_Full and the MIL_100_TTS_VC were further analyzed for TTC subway stations that have boardings or alightings that exceed 10,000 according to the TTS counts. The Bloor–Yonge Station (large bars in the middle of Figures 6*a* and 6*b*), a major transfer point between the two subway lines, has 1,736 boardings and 12,628 alightings in the a.m. peak period. For this station, EMME/2 overestimates boardings and alightings by 12.71% and 11.38%, respectively; whereas MILATRAS overestimates boardings and alightings for this major subway station by 6.62% and 1.36%, respectively.

UNIQUENESS OF MILATRAS

Unlike other transit assignment modules, MILATRAS uses an agent-based mesoscopic representation in a microsimulation framework for modeling passenger path choice and behavior and for modeling transit system characteristics. This gives MILATRAS unique capabilities for modeling emerging smart transit systems, such as analyzing the effects of vehicle capacity, providing traveler information, and presenting detailed passenger information.

Vehicle Capacity

In EMME/2 and MADITUC, special consideration is required to include vehicle capacity constraints. However, the microsimulation-based framework in MILATRAS has capacity constraints built into the network. Incorporating vehicle capacity improves model realism because passengers cannot board a full vehicle and need to wait for the next arriving vehicle. This constraint is important particularly during congested times of day, in particular during the a.m. peak period.

To illustrate the importance of vehicle capacity constraints, Route 53-B Steeles East, with a bus capacity of 51 passengers, was examined. MADITUC_SIM shows this route to have a maximum load factor of 1.63; the maximum load factor is defined as the maximum number of passengers divided by the theoretical capacity. EMME/2_V3_Dis_Full shows Route 53-B to have a maximum load factor of 6.6. Both MADITC and EMME/2 have maximum load factors that are greater than 1. This assumes that all passengers can board the next arriving bus, which does not reflect reality. MILATRAS_100_TTS_VC incorporates vehicle capacity and has a maximum load factor of 1.00 (removing the vehicle capacity gives Route 53-B a maximum load factor of 4.35). The importance of the vehicle capacity constraint can be seen in modeling actual demand for Route 53-B. MADITUC overestimated Route 53 by 1.14%; this is in contrast to the rest of MADITUC's modeled routes that are within 1% of modeled loads. EMME/2 overestimated Route 53 by 1.78%. MILATRAS, with vehicle capacity constraints, produced outputs similar to actual loads.

Traveler Information

With the rapid growth of ITS applications, the need for dynamic models of travel behavior and network performance has been growing. The provision of real-time travel information on transit services is increasingly being recognized as a potential strategy for influencing transit rider behavior on departure time choice and path choice and, it is hoped, attracting auto-mode users. Understanding travelers' responses to this information is therefore critical to the design and implementation of effective intelligent transport systems strategies such as ATIS. The benefits of ATIS applications can be assessed by comparing the path choice behavior and time savings of informed passengers with noninformed passengers. However, conventional transit assignment models, such as EMME/2 and MADITUC, assume that passengers have full information about the network conditions and infinite information processing capabilities; this is referred to as the "perfect knowledge of network" assumption. These models are not appropriate for modeling information provision because information on network conditions is assumed to be available, anyway, to all passengers. The emergence and increased deployment of ATIS make it practically important to relax the assumption of perfect information or perfect knowledge in transit assignment studies. Wahba and Shalaby show how MILATRAS can be used to investigate the impact of various scenarios of information provision (*16*).

More transit systems today are incorporating advanced technological systems to improve operational efficiency and customer satisfaction. TTC has recently incorporated ATIS at major subway stations, including Union Station and Spadina Station, to show passengers when the next train or streetcar is arriving. The agent-based microsimulation framework of MILATRAS allows for proper representation of a transit network that includes such technologies.

Detailed Passenger Information

The agent-based representation of passengers in MILATRAS allows planners to extract detailed passenger and route information unavailable through EMME/2 and MADITUC. In contrast to the logit-based model in EMME/2 and MADITUC, the learning-based framework of MILATRAS models passengers' experiences and choices explicitly. In both the aggregate and disaggregate transit assignment procedures in EMME/2, individual information is lost through aggregation. An agent-based implementation preserves and exploits the disaggregate information available through the TTS survey.

Wahba and Shalaby show an example of the choice probabilities for a passenger–agent in comparison with the observed choices (*14*). In MILATRAS, the mental model (or choice set) contains information on all other options available for passenger–agents, and not chosen. This is a unique feature for the proposed framework because not only does it explain why a certain travel option is chosen by a passenger, but it also provides information on why a certain travel option is not chosen. This is important for analyzing passengers' preferential treatment of travel options. In equilibrium-based models, passengers do not change their choices unilaterally; the choices of other passengers are required for the decision-making process. In reality, passengers change their choices on the basis of their perception of the travel cost of available options; when the service characteristics of these options change (e.g., faster bus service), passengers may change their trip choices without knowing others' decisions. Through learning and adaptation within the microsimulation environment, a new steady state of passengers' choices and service performance can be reached.

Other unique features of MILATRAS' structure include the combination of the departure time choice, stop choice, and run (or sequence of runs) choice in one framework along with the representation of day-to-day and within-day dynamics in travelers' choices as well as transit service. MILATRAS is unique in dealing with the network-level effects of such interactions; existing approaches deal with fewer choice dimensions (e.g., ignore stop choice) on the network level.

CONCLUSIONS AND FUTURE RESEARCH

MILATRAS presents a new learning-based approach to the transit assignment problem that dynamically models transit service characteristics and passenger behavior and experience and represents the interaction between the network supply and travel demand explicitly. The agent-based microsimulation framework allows MILATRAS to accurately represent today's evolving public transit service.

This paper presents a model validation of MILATRAS with two other common transit assignment packages, EMME/2 and MADITUC, by using the TTC service network. This large-scale application represents about 500 branches with more than 10,000 stops and about 332,000 passengers for TTC in the morning peak period. Although MILATRAS presents a different standpoint for analysis in comparison with equilibrium-based models, it still gives the steady state run loads (comparable with stochastic user equilibrium loads). MILATRAS presents a stochastic process approach (i.e., nonequilibrium based) for modeling within-day and day-to-day variations in the transit assignment problem in which aggregate travel patterns can be extracted from individual choices. As shown, with the TTC network, the predictive performance of MILATRAS was very promising, performing as well as EMME/2 and MADITUC. MILATRAS, in addition, presents a policy-sensitive platform for modeling the effects of smart transit system policies and technologies on passengers' travel behavior (i.e., trip choices) and transit service performance.

Recently, MILATRAS has been integrated within a multimodel framework to evaluate transit emissions on a link-based mesoscopic level for the TTC network by using time-dependent speed and loading profiles produced by MILATRAS (*17*). Future efforts will be directed to incorporating access and egress mode choices for the modeling of multimodal trips. Moreover, a mode choice (or a mode shift) component can be added to the trip choice hierarchy. The integration of dynamic transit assignment models (such as MILATRAS) and activity-based urban planning models is needed because transit assignment is a key component of land use and transportation models. Eventually, MILATRAS is envisioned to be used in designing transit networks and services.

REFERENCES

1. Wahba, M., and A. Shalaby. MILATRAS: A New Modelling Framework for the Transit Assignment Problem. In *Schedule-Based Modeling of Transportation Networks: Theory and Applications* (N. Wilson and A. Nuzzolo, eds.), Springer-Verlag, Berlin, 2009, pp. 171–194.
2. Wahba, M., and A. Shalaby. Learning-Based Departure Time and Path Choice Modelling for Transit Assignment under Information Provision: A Theoretical Framework. Presented at 12th International Conference on Travel Behaviour Research, Jaipur, Rajasthan, India, Dec. 13–18, 2009.
3. Wahba, M. MILATRAS: Microsimulation Learning-Based Approach to Transit Assignment. PhD dissertation. Graduate Department of Civil Engineering, University of Toronto, Ontario, Canada, 2008.
4. MADITUC. Toronto, Ontario, Canada, 2008. http://www.transport.polymtl.ca. Accessed Feb. 20, 2009.
5. Miller, E. J. A Travel Demand Modeling System for the Greater Toronto Area, Version 3.0. Joint Program in Transportation, University of Toronto, Ontario, Canada, 2007.
6. Wahba, M., and A. Shalaby. *Transit Assignment Approaches: Past, Present and Future.* Working paper. Civil Engineering Department, University of Toronto, Ontario, Canada, 2009.
7. Chriqui, C., and P. Robillard. Common Bus Lines. *Transportation Science,* Vol. 9, 1975, pp. 115–121.
8. Le Clercq, F. A Public Transport Assignment Method. *Traffic Engineering and Control,* Vol. 17, No. 1, 1972, pp. 14–20.
9. Spiess, H., and M. Florian. *Optimal Strategies: A New Assignment Model for Transit Networks.* Centre de recherché sur les transport, Université de Montreal, Quebec, Canada, 1989.
10. *EMME/2 User's Manual,* Release 9.4. INRO, Montreal, Quebec, Canada, 2003.
11. Chapleau, R. Transit Network Analysis and Evaluation with a Totally Disaggregate Approach. Presented at World Conference on Transportation Research, Vancouver, British Columbia, Canada, 1986.
12. Nuzzolo, A., F. Russo, and U. Crisalli. Transit Network Modelling: The Schedule-Based Approach. *Collana Trasporti,* FrancoAngeli s.r.l., Milan, Italy, 2003.
13. *Transportation Tomorrow Survey 2001: Design and Conduct of the Survey.* Data Management Group, University of Toronto, Joint Program in Transportation, Ontario, Canada, Jan. 2003.
14. Wahba, M., and A. Shalaby. Large-Scale Application of MILATRAS: Case Study of the Toronto Transit Network. Presented at 12th International Conference on Travel Behaviour Research, Jaipur, Rajasthan, India, Dec. 13–18, 2009.
15. *Transportation Tomorrow Survey 2001: Data Validation.* Data Management Group, University of Toronto, Joint Program in Transportation, Ontario, Canada, Feb. 2003.
16. Wahba, M., and A. Shalaby. An Agent-Based Microsimulation Model with Experiential Learning for Passenger Assignment under Information Provision. Accepted for presentation at 11th International Conference on Advanced Systems for Public Transport, July 20–22, 2009, Hong Kong (not presented).
17. Lau, J., M. Hatzopoulou, M. Wahba, and E. Miller. *An Integrated Multi-Model Evaluation of Transit Bus Emissions in Toronto.* Working paper. Civil Engineering Department, University of Toronto, Ontario, Canada, 2009.

The Transportation Demand Forecasting Committee peer-reviewed this paper.

Validation and Forecasts in Models Estimated from Multiday Travel Survey

Elisabetta Cherchi and Cinzia Cirillo

Multiday travel surveys (also known as panel data) have recently assumed high relevance in travel behavior analysis and activity-based modeling. Basically two types of multiday data have been collected: cross-sectional data repeated at separate points in time and data gathered over a continuous period of time. So far, the studies using panel data have focused on the estimation of demand models; little is known about the application and validation of models estimated on repeated measurements. Even the definition of the holdout sample is not obvious in panel data sets. This paper studies issues related to model validation and forecasting of continuous data sets. With both simulated and real data, empirical evidence is provided on the effects that different patterns of correlation have on model forecast and policy analysis. Results show that the way holdout samples are extracted affects the validation results and that the best results are obtained when a percentage of individuals with all of their observations are used. The logit model in the presence of taste heterogeneity could produce biased modal shifts, while failing to account for correlation across observations, did not seem to produce relevant effects on policy analysis. The real case study, estimated by using a 6-week travel diary (MobiDrive), confirms only in part the analysis on simulated data. Results also confirm that in panel data, a model with a better fit might provide a worse validation and forecast.

A number of multiday travel surveys (also generally called panel data) have been collected in the past decade, and consequently their use for travel behavior modeling and analysis has significantly increased. There are basically two types of multiday data depending on whether the same survey is repeated at "separate" times (e.g., once or twice a year for a certain number of years), or over a "continuous" period of time (e.g., 7 or more successive days).

The first type of survey has been used to gain insights into activity scheduling and travel planning, to study dynamic effects such as habit and learning, and in general to study how behaviors change as the environment varies (i.e., the supply or the socioeconomic characteristics). Golob used three waves of data (1 year apart) from the Dutch National Mobility Panel to study inertial and lagged relationships between income, car ownership, and car and public transportation use (*1*). Bradley, with before and after data, estimated dynamic logit models that account for response lags and state dependence to study the effect on mode choice of a new rail commuter line (*2*). Simma and Axhausen with the use of panel data from both Germany and the Netherlands found that travel commitments (car ownership and public transport season tickets) in one period affect mode usage in the next period (*3*). Chatterjee and Ma used a panel of four waves to examine the time scale of behavioral responses in travel mode shifts when change tends to take longer to occur (*4*). Thørgersen used structural equations modeling and three waves of travel data (between 1998 and 2000) to study to what extent the current behavior toward public transport is influenced by past behavior, current attitudes, and perceived behavioral control (*5*). Srinivasan and Bhargavi used a data set recorded on two time points during a period of 5 years to account for rapid and substantial changes in the fast-growing Indian economy (*6*). Ramadurai and Srinivasan used a set of data gathered on 2 consecutive days to study habit persistence and state dependence in modal choice (*7*). Finally, Yañez and Ortúzar studied the effect of shock and inertia in individual behavior by using a 5-day pseudodiary that has been repeated four different times so far, just before and three times after the implementation of the radically new and much maligned Santiago public transportation system (Transantiago) in Chile (*8*).

Numerous recent studies about travel behavior have been based on continuous panel data. Multiday travel diaries have been used to detect rhythms of daily life (*9*), to compare different indices that measure similarities of travel behavior (*10*), and to analyze the variability in daily travel of individuals and the proportions of variance arising from intrapersonal and interpersonal variability (*11*). Advanced econometric models have been applied to continuous panels to draw evidence on the parametric assumptions behind the value of time distribution (*12*), to study day-to-day variability in modal choice models (*13*), and to examine the length of time between successive participations in several activity purposes (*14*). Recently, researchers are attempting to introduce dynamics into activity-based model systems; demand model systems for daily activity programming based on a 1-week travel diary have been estimated (*15*), and an extended reinforcement learning approach has been proposed to produce weekly activity patterns in Belgium (*16*). Arentze and Timmermans have used the need-based framework for defining dynamic activity utility functions and to develop a heuristic method to generate activity agendas on a multiday, multiperson basis (*17*). Finally, discrete choice models estimated with panel data have been proposed to explain current behavior on the basis of individuals' history and experience (*18*), to examine the effect of repeated observations (*19*), and to try to account for two different correlation effects across individuals over two time periods along the panel (*20*).

This literature review is not exhaustive by any means, but it clearly shows the strong interest in multiday data sets and the major advancements that have been produced so far by the transportation community. Nevertheless, the main focus of almost all of these studies is the

E. Cherchi, Research Center in Mobility Patterns, Department of Engineering, University of Cagliari, Piazza d'Armi 16, 09123 Cagliari, Italy. C. Cirillo, Department of Civil and Environmental Engineering, University of Maryland, 1179 Glenn Hall, College Park, MD 20749. Corresponding author: E. Cherchi, echerchi@unica.it.

Transportation Research Record: Journal of the Transportation Research Board, No. 2175, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 57–64.
DOI: 10.3141/2175-07

estimation of demand models (mainly mode choice and activity participation), although their prediction power is still not well known. Moreover, related to that, the validation of models estimated on panel data has never been studied in depth. Because the ultimate objective of transport modeling is demand forecasting, the authors believe that it is crucial to explore to what extent models estimated on multiday data can be used for prediction purposes.

Prediction and validation issues have been explored up to a certain level in panel data gathered on separate points in time; the results of model estimation have actually been used to study how (travel) demand varies when external changes occur at one point in time. The inclusion of past behaviors, lagged responses, and state-dependence effects has in fact proved to improve model fit and to affect policy analysis. However, models estimated on continuous multiday data aim mainly at investing intrinsic day-to-day variability, but no major external changes usually occur along the period of the survey collection. Hence the problem in prediction arises: How can models estimated on continuous panels be used? And more precisely, what time frame should these models apply to? As discussed by Cherchi et al., properly accounting for correlation over repeated observations is crucial to understand the long-term structure of individual choices (*20*). It is indeed crucial not only to produce correct estimates, but also for their validation and use in forecasting. But this issue has never been explored, and the application and validation of models estimated on repeated measures is still an unsolved problem.

In the case of validation, another problem arises because not even the definition of the holdout sample is obvious in panel data sets. Holdout samples are generally drawn randomly from cross-sectional data; in multiday data different validation sets can be formed depending on the dimension that the analyst has decided to adopt. In fact, having at least two dimensions of variability, individuals and their answers, two types of holdout samples can be drawn: (*a*) a subsample of individuals, each with the full set of answers, or (*b*) a full set of individuals, with only a subset of answers from each individual.

The objective of this paper is to explore the use of models estimated on continuous panel data for validation and forecasting purposes. By using simulated and real data, on the basis of the 6-week panel gathered in Germany, empirical evidence will be provided on (*a*) the effects on prediction and validation of different types of correlation and (*b*) the effects on validation of different ways to draw holdout samples.

The rest of the paper is organized as follows. In the next section discrete choice models for panel data are formalized and the different dimensions of correlations in multiday–multiweek data sets are introduced. Then the analysis on simulated data is reported, in which perfect correlation is assumed for responses from the same individual. This hypothesis is consistent with the model formulation assumptions. Validation and prediction issues for a real case study on a 6-week travel survey are the object of the next section. Principal findings, lessons learned, and avenues for future research conclude the paper.

MODEL FORMULATION

A mixed logit formulation to model mode choice on panel data is adopted (*21*). It is assumed that for a single tour in each choice situation the person chooses among a finite set of alternatives; the choice set can vary over tour episodes and the number of choice situations can vary over days, weeks, individuals, and households. The utility that person q obtains from alternative j in each choice situation is as follows:

$$U_{qj}^{t} = \sum_{k=1,\ldots,K} \left(b_k + \mu_{qk} + b_k^{SE} SE_q\right) X_{kqj}^{t} + \sum\nolimits_m \left(b_m + \mu_{qm}\right) y_{jm} + \epsilon_{qj}^{t} \quad (1)$$

where

X_{kqj}^{t} = level of service variables that vary among individuals q, alternatives j, and over time periods t;
SE = socioeconomic variables;
b_k and b_m = parameters fixed over population and time periods;
μ_{qk} and μ_{qm} = individual parameters fixed over time periods and randomly distributed with zero mean;
y_{jm} = index that equals 1 if m appears in utility function j, 0 otherwise, allowing for error component; and
ϵ_{qj}^{t} = Gumbel distributed random terms.

The specification in Equation 1 is quite generic as it allows accounting for systematic and random heterogeneity around the responses to level-of-service attributes, different forms of correlation among alternatives, random heterogeneity in the preferences for specific alternatives, and correlation across tour mode choices made during the same day, the same week, or by the same individual. In cases of more than one observation available for each individual, given the sequence of modal alternatives, one for each tour episode $j = \{j_1, \ldots, j_T\}$, the probability that the person makes this sequence of choices is the product of logit formulas (*22*), where V is the utility as in Equation 1, excluding the Gumbel error:

$$L_{qj}(b,\mu) = \prod_{t=1}^{T} \left[\frac{e^{V_{qj}^{t}(b,\mu)}}{\sum_l e^{V_{qj}^{t}(b,\mu)}} \right] \quad (2)$$

If the number of choices in each sequence equals 1 ($T = 1$), then the specification degenerates to mixed logit on cross-sectional data; if T is greater than 1, the formulation allows correlation across the observations belonging to the same sequence.

The unconditional probability is the integral of this product over all values of μ:

$$P_{qj} = \int L_{qj}(b,\mu) f(\mu) d\mu \quad (3)$$

When using cross-sectional data in which the person's previous choices are not known, one mixes the logit formula over density of μ in the entire population. However, when the person's previous choices are known, it is expected that one can improve the prediction by mixing over the density of μ in the sequence (*21*).

The vector of unknown parameters is then estimated by maximizing the log likelihood (LL) function, that is, by solving the equation as follows:

$$\max_{b,\mu} LL(b,\mu) = \max_{b,\mu} \sum_{1}^{Q} \ln P_{qj}(b,\mu) \quad (4)$$

where j_t is the alternative choice made by individual q in time period t. This involves the computation of $P_{qj}(b, \mu)$ for each individual q, $q = 1, \ldots, Q$, which is impractical because it requires the evaluation of one multidimensional integral per individual. The value of $P_{qj}(b,\mu)$ is therefore replaced by a Monte Carlo estimate (SP) obtained by sampling over μ, and given by

$$SP_{qj}^{R} = \frac{1}{R} \sum_{r=1}^{R} P_{qj_t}(b,\mu) \quad (5)$$

where R is the number of random draws μ_r, taken from a predefined parametric distribution. As a result, b and μ are now computed as the solution of the simulated log likelihood (SLL) problem:

$$\max_{b,\mu} \mathrm{SLL}^R(b,\mu) = \max_{b,\mu} \frac{1}{Q}\sum_{q=1}^{Q} \ln \mathrm{SP}_{qj}^R(b,\mu) \quad (6)$$

The solution of this last approximation [often called sample average approximation (SAA)] will be denoted by a $(b, \mu_R)^*$; $(b, \mu)^*$ denotes the solution of the true problem (Equation 4). All the results presented in this paper are obtained with the software AMLET (www.grt.be/amlet), which uses an adaptive stochastic programming algorithm to estimate mixed logit models (*23*).

EXPERIMENTS USING SIMULATED DATA

Repeated observations from synthetic individuals have been simulated to study validation issues and policy-related effects in discrete choice models. Simulated data are necessary to ensure that the model to be estimated is coherent with the hypothesis formulated by the analysts and reported in the previous section. The synthetic sample is composed of 200 individuals, each of them is supposed to provide 20 valid responses; a total of 4,000 observations are then generated. The number of repeated observations and the number of alternatives were chosen to generate a synthetic population with characteristics similar to that of the real panel data. The 20 observations provided by each individual can be thought of as if each person described two trips or tours a day during a 2-week period, with 5 working days in each week.

In the experimental context respondents choose between five alternatives; random utilities are specified with a full set of alternative specific constants (ASC1, ASC2, ASC3, and ASC4) assumed to be constant across individuals and two generic level-of-service coefficients normally distributed [time (t) and cost (c)]. Details of the sample characteristics are given in Table 1. The sample was generated accounting for correlation among the 20 responses of each individual, as normally found in real data. Following the discussion in the section on model formulation, correlation was accounted for in the random coefficients.

TABLE 1 Simulated Data: Utility Specification

Alternative	Constant	Coefficients	Variables
Alt1	Time (t)	N(–0.05, 0.05)	N(20, 20)
	Cost (c)	N(–0.5, 0.5)	N(2.0, 3.0)
Alt2	ASC1	–0.5	1
	Time (t)	N(–0.05, 0.05)	N(20, 20)
Alt3	ASC2	–1.5	1
	Time (t)	N(–0.05, 0.05)	N(40, 15)
	Cost (c)	N(–0.5, 0.5)	N(1.0, 2.0)
Alt4	ASC3	–0.8	1
	Time (t)	N(–0.05, 0.05)	N(30, 30)
Alt5	ASC4	0.3	1
	Time (t)	N(–0.05, 0.05)	N(20, 20)

NOTE: ASC = alternative specific constant.

With the simulated data above, four different model formulations were estimated: (a) multinomial logit (MNL), (b) mixed logit (ML) on cross-sectional data (ML-cross), (c) ML on panel data with correlation over two observations (ML-panel2), and (d) ML on panel data with correlation over 20 observations (ML-panel20). Moreover, all of these models were estimated on a subsample of 3,200 observations to study the validation of mode choices by using panel data. In particular, following the discussion in the previous section, the holdout sample was generated in two different ways: a first sample was created including all observations belonging to the first 160 individuals, that is, keeping the remaining 20% of the sample out for validation purposes, and a second sample was created including all 200 individuals but only the first 16 answers of each individual, that is, leaving the last four answers of each individual out for validation. Table 2 shows the results from models estimated with the first sample, that is, 160 individuals but with 20 answers each.

As expected the fit of the model improves gradually but significantly when heterogeneity on time and cost is accounted for and when correlation effects are considered. In particular, when correlation is applied to the full set of responses given by the same individual, the adjusted rho-squared increases from 0.265 to 0.345. In addition, although all estimated parameters are highly significant, only the ML

TABLE 2 Simulated Data: Model Estimation

		MNL		ML-Cross		ML-Panel2		ML-Panel20	
Alternative	True Values	Estim.	t-Stat.	Estim.	t-Stat.	Estim.	t-Stat.	Estim.	t-Stat.
ASC1	–0.5	–0.5737	9.1	–0.4948	6.0	–0.5103	6.6	–0.4808	22.5
ASC2	–1.5	–1.2821	12.8	–1.4586	12.2	–1.4959	13.0	–1.4976	42.9
ASC3	–0.8	–0.7544	11.1	–0.7728	8.0	–0.7992	9.4	–0.8081	29.5
ASC4	0.3	0.1636	3.0	0.3541	4.4	0.3379	5.1	0.3654	8.2
Time (mean)	–0.05	–0.0330	26.0	–0.0478	19.6	–0.0486	27.3	–0.0524	6.8
Time (SD)	0.05	—	—	0.0412	12.1	–0.0428	17.2	0.0473	9.0
Cost (mean)	–0.5	–0.2770	15.0	–0.4819	13.2	–0.4814	16.0	–0.5043	6.2
Cost (SD)	0.5	—	—	0.4738	8.3	0.4732	13.1	0.5208	9.6
N observed		3,200		3,200		3,200		3,200	
N individuals		160		160		160		160	
LL (0)		–5,150.20		–5,150.20		–5,150.20		–5,150.20	
LL (final)		–3,933.31		–3,861.73		–3,785.16		–3,373.96	
Adj. rho-squared		0.235		0.250		0.265		0.345	
CV (scale parameters)		0.31		0.10		0.08		0.08	

NOTE: N = number of samples, LL = log likelihood, CV = coefficient of variation, and SD = standard deviation.

models are able to recover the true values. The last row of Table 2 reports the coefficient of variation (CV) of the ratio between the estimated and the true parameters, which is the scale of the estimated model (*24*). If the estimated parameters differ only from the true values, then this ratio is equal among all the parameters inside each model. The bigger the CV, the poorer the model's ability to reproduce the true phenomenon. Note also that accounting correctly for the correlation improves model fit but not the model's capability to recover the true phenomenon. This result confirms previous findings that the effect of the correlation is (or should be) only that of allowing lower estimated variance (*13, 19*).

Coefficient estimates are then used in application to calculate the prediction power of the models; modal shares observed and predicted together with a measure of the errors are reported in Table 3. Two error measurements are used, the absolute error and the 2-norms of the distance between predicted and observed modal shares. The absolute error D is defined as

$$D = \left| M_{\text{pred}} - M_{\text{obs}} \right| \tag{7}$$

where

D = error norm,
M_{pred} = vector of modal shares predicted, and
M_{obs} = vector of modal shares observed.

The 2-norms is defined as

$$\text{2-norms} = \sqrt{D_{\text{Alt1}}^2 + D_{\text{Alt2}}^2 + D_{\text{Alt3}}^2 + D_{\text{Alt4}}^2 + D_{\text{Alt5}}^2} \tag{8}$$

where $D_{\text{Alt1}}, \ldots, D_{\text{Alt5}}$ are the components of the error vector.

The analysis of the results obtained with the model estimated on 160 individuals and 20 observations per individual (Table 3) indicates that slight improvements are registered in the error norm and 2-norms when the model is correctly estimated with heterogeneity in tastes (ML-cross), heterogeneity and correlation over two observations (ML-panel2), and heterogeneity and correlation over 20 observations (ML-panel20), which corresponds to the "true" model. However, if the total market share predicted is considered, it may be concluded that all models, including the MNL, perform quite well and that the correct one is the best (as expected).

When models were estimated with the second sample, that is, all 200 individuals but only the first 16 answers, it was found that all models perform quite similarly to what is discussed in Table 2, except model ML-panel2, which surprisingly failed to converge properly. Table 3 reports the results of the model validation. As a consequence of the poor estimation, the model ML-panel2 shows the biggest errors in predictions (slightly higher errors), whereas the remaining models estimated perform equally well in predictions.

Finally, in comparing the errors in Table 3 it is important to note that the error is smaller in magnitude when the model is estimated on the full set of respondents and validated on part of the individual responses. Another important result is that in both experimental cases the specification on panel data (which is assumed to be the correct one) produces a better fit, but does not improve the prediction power of the model.

The policy analysis in models estimated on data with repeated observations is now considered. Only the policy analysis using the results of the models estimated with 160 individuals and 20 responses each will be reported. Using the model estimated with 200 individuals and 16 observations each, produces very similar results. The effects on modal shares of a 50% increase in time for Alternative 1 and Alternative 2 and a 50% increase in cost for Alternative 1 are studied. Table 4 reports the changes in the aggregate share of mode j over the initial situation (do nothing):

$$\Delta P_j = \frac{P_j - P_j^0}{P_j^0} \tag{9}$$

TABLE 3 Model Validation: Modal Shares

		Predicted			
Alternative	Observed	Logit	ML-Cross	ML-Panel2	ML-Panel20
Sample: 160 Individuals—20 Observations per Individual					
Alt1	205	188	189	191	191
Alt2	168	164	163	163	165
Alt3	44	36	37	36	36
Alt4	103	109	108	108	107
Alt5	280	303	303	302	301
D	—	58	57	55	50
2-norms	—	30.7	30.0	28.6	26.7
Sample[a]: 200 Individuals—16 Observations per Individual					
Alt1	180	189	187	150	186
Alt2	158	167	168	188	169
Alt3	35	36	36	34	39
Alt4	121	109	109	120	110
Alt5	306	299	299	307	297
D	—	38	38	62	42
2-norms	—	18.9	18.7	42.0	19.7

NOTE: D = error norm.
[a]Model ML-Panel20 did not achieve final converge.

TABLE 4 Policy Analysis

	+50% Time Alt. 1 and Alt. 2; +50% Cost Alt. 1				
Alternative	True	Logit	ML-Cross	ML-Panel2	ML-Panel20
Sample: 160 Individuals—20 Observations per Individual					
Alt1	−0.072	−0.102	−0.080	−0.080	−0.078
Alt2	−0.058	−0.091	−0.069	−0.067	−0.066
Alt3	0.034	0.086	0.052	0.050	0.047
Alt4	0.039	0.071	0.049	0.048	0.046
Alt5	0.057	0.072	0.060	0.060	0.060
D	—	0.162	0.051	0.045	0.039
2-norms	—	0.077	0.025	0.022	0.019
Sample: 200 Individuals—16 Observations per Individual					
Alt1	−0.109	−0.103	−0.100	−0.106	−0.108
Alt2	0.095	0.151	0.110	0.109	0.095
Alt3	1.189	1.092	1.279	1.314	1.313
Alt4	0.475	0.478	0.530	0.529	0.538
Alt5	−0.319	−0.317	−0.338	−0.336	−0.335
D	—	0.165	0.189	0.212	0.204
2-norms	—	0.113	0.109	0.138	0.139

where P_j^0 and P_j are, respectively, the aggregate probabilities of choosing mode j before (do nothing) and after introducing the measure, calculated by sample enumeration (*25*). In particular, the first column reports the simulated market share variation, that is, computed directly from the simulated data set, and the remaining columns report the market share variation computed by applying the different estimated models.

The interpretation is quite straightforward; the multinomial logit fails to predict the true modal shifts when variation in travel time is applied. When heterogeneity is correctly accounted for with random normally distributed coefficients, the error norm and the 2-norms indicator decrease in the +50% time scenario and slightly increase in the 50% cost scenario. Models accounting for correlation effects over two and 20 observations continue to improve model predictions in the +50% time scenario but leave probabilities in the other scenario almost unchanged. In conclusion, the logit model in the presence of taste heterogeneity can produce biased modal shifts, while failing to account for correlation across observations does not seem to produce relevant effects on policy analysis.

EVIDENCE FROM A REAL CASE

In this paper observations on mode choice are extracted from the 6-week travel diary collected in Karlsruhe, Germany, and part of the MobiDrive survey. The original study involved about 160 households and 360 individuals in the cities of Karlsruhe and Halle, Germany (*26, 27*). The derived data set includes mode choice observations at the tour levels; all tours on a daily basis are considered. Daily patterns are derived by applying the framework proposed by Bhat and Singh in 2000 (*28*). A tour is the sequence of trips performed by an individual, starting from a given base (usually home or workplace) until the individual returns to this base. Each tour has a main activity defined by duration, purpose, and main mode. The main activity of the day is assumed to be work or education. The work tour is divided into outbound and return legs, which are called the morning and evening commute. All activities that take place before the morning commute will be referred to as morning activities and the associated displacements will be grouped into one or more morning tours; they constitute the morning pattern. Similarly, all activities taking place after the return from work to home (the evening commute) will be referred to as evening activities and the associated displacements grouped into one or more evening tours; which together constitute the evening pattern. In addition, all activities taking place outside the work location after the morning, but before the evening commute, will be called midday activities, and the associated displacements, whose origin and destination are at work, are grouped into one or more midday tours, in turn aggregated into the midday pattern. In this paper, only working days from Mondays to Fridays are considered; preliminary analysis indicated that modal choices registered during the weekend are substantially different from weekday shares and deserve separate investigation. In summary, after the described framework is applied and the aforementioned exclusions are considered, 4,089 activity episodes, 2,488 daily schedules, 674 weekly schedules, 126 individual schedules, and 56 household schedules are obtained.

The specification used to test model validation and policy analysis together with the results from the whole sample is shown in Table 5. The model presents 24 degrees of freedom; the coefficients multiplying the two level-of-service variables (time and cost) are normally distributed; the remaining coefficients including the four alternative specific constants are all constant. Deterministic variability in travel time is captured by the interaction terms with four different types of socioeconomic variables. The remaining variables are specific of the alternatives and include sociodemographic and location characteristics (age, household location, marital status) and activity episode attributes (time budget, purposes).

Similar to what is already done in the simulated experiments, by using MobiDrive data the following are estimated: MNL, ML-cross, ML-panel_d, and ML with correlation over the entire sets of responses from the same individual (ML-panel_i). As expected, the fit of the model improves when heterogeneity and correlation effects are included in the model formulation; the model allowing for heterogeneity in level-of-service variables and accounting for correlation over individual observations provides the best fit.

For validation purposes, models have been reestimated on a subsample. Results are similar to those illustrated in Table 5 and are not reported. Given the different (temporal) dimensions of a multiday/multiweek travel survey, two model validations have been carried out: (*a*) the model has been estimated on the observations from the first 100 individuals and validated on the tours from the remaining 26 individuals and (*b*) the model has been estimated on the entire set of individuals and validated on the 6th week from each individual. The idea is that by conditioning individual probabilities on the first 5 weeks, the model prediction power on the 6th week should improve.

Table 6 shows the observed versus predicted market shares calculated from the model estimated on 100 individuals; here the errors in predictions increase with the number of effects considered (heterogeneity, correlation across daily tours, and correlation across individual tours). Different from the simulated data, with real data what the right specification is, is not known, but it is interesting that the model with the best fit does not provide the best validation. This result seems also to confirm that accounting for correlation improves the fit but not the model's capability to reproduce the true phenomenon. The result does not change when validating the model on the 6th week (Table 6). The best prediction is provided by the MNL model, although differences (in absolute error and 2-norms indicator) with the ML model not accounting for correlation (ML-cross) are small. The model prediction power rapidly deteriorates when day and individual correlations are introduced. These results suggest that the best statistical fit does not guarantee that the model is the best in reproducing the real phenomenon. These results extend to the panel data results found in Cherchi and Ortúzar (*29*).

Consistently with what is observed on simulated data experiments, the error norm and 2-norms are smaller in magnitude when individual mode choices are conditioned on the first 5 weeks. In both cases considering correlation effects in model estimation does not improve the model performance in predictions.

Two different policy scenarios are analyzed for MobiDrive. In the first case modal shifts are calculated when an increase of 30% in car travel time (for both drivers and passengers) is applied (Table 7), and in the second case car driver cost is expected to increase 50% (Table 7). Both cases have been applied to the model estimated on the first 100 individuals with their entire sets of observations. In a real case it cannot be discerned which are the "true" forecasts, or those closer to the actual individual behavior. In the +30 time scenario, modal shifts are similar for logit and ML-cross; whereas differences in choice probabilities before and after the policy is applied gradually increase. For the +50% cost scenario it is not easy to find a pattern in model predictions when heterogeneity in time and cost and different structure of correlation are considered. If one believes in the statistical results, the model with the best fit (ML-panel_i)

TABLE 5 MobiDrive Model Results

Alternative[a]	Coefficient	Logit (MNL) Estim.	Logit (MNL) *t*-Stat.	Mixed (ML-cross) Estim.	Mixed (ML-cross) *t*-Stat.	Mixed Day (ML-panel_d) Estim.	Mixed Day (ML-panel_d) *t*-Stat.	Mixed Indiv. (ML-panel_i) Estim.	Mixed Indiv. (ML-panel_i) *t*-Stat.
Alternative Specific Constants (car driver is the base)									
CP	ASC CP	−0.3804	1.0	−0.5953	3.2	−0.7714	2.0	−0.1969	1.1
PT	ASC PT	0.4795	1.4	0.7262	3.2	1.1093	3.1	0.8362	4.0
W	ASC W	−0.7175	2.0	−1.0127	6.4	−1.1214	3.0	−0.3220	1.7
B	ASC B	−1.8544	4.9	−2.2433	11.9	−2.5406	6.7	−1.9835	10.9
Sociodemographic and Location Attributes as Alternative Specific									
CD	Main car user	2.4018	2.5	2.2683	5.8	2.5991	4.3	2.7461	12.6
PT	Urban location	0.2138	1.1	0.1055	0.8	0.1457	0.7	0.7560	3.1
PT	Married with child(ren)	0.8335	4.9	0.9284	7.4	1.1131	5.0	1.4857	15.4
CD	Age 26–35 years	2.0274	3.2	1.9367	5.4	2.0539	4.4	2.5223	8.8
PT	Age 51–65 years	0.6401	3.8	0.7335	5.9	0.8610	4.2	0.6981	5.2
W	Age 51–65 years	1.9179	11.0	1.8842	16.1	1.9957	11.7	2.1706	15.1
B	Age 18–25 years	0.1123	0.5	0.0773	0.6	−0.0093	0.0	−0.1852	2.0
B	Age 26–35 years	1.4862	4.4	1.5720	7.8	1.6855	7.4	1.5748	22.8
B	Age 51–65 years	0.8335	4.9	0.9284	7.4	1.1131	5.0	1.4857	15.4
Activity Episode Attributes									
CP	Leisure	−0.3007	0.9	−0.3255	1.9	−0.3474	1.9	−0.2427	3.0
PT	Leisure	1.1196	6.3	1.2638	8.4	1.3604	6.7	0.8543	9.6
PT	Work	0.0006	1.7	0.0007	3.0	0.0007	1.9	0.0006	3.4
CD	Time budget	0.0010	3.7	0.0011	6.0	0.0012	5.1	0.0012	7.8
B	Time budget	−0.0129	3.4	−0.0139	3.5	−0.0219	3.7	−0.0321	3.2
Interactions Time * Sociodemographic									
All	Time * married with child(ren)	−0.0042	0.8	−0.0054	0.9	−0.0064	0.9	−0.0046	0.3
All	Time * female and part-time	−0.0234	5.1	−0.0280	5.5	−0.0299	5.3	−0.0061	1.6
All	Time * number of stops	0.0194	5.9	0.0196	5.6	0.0279	5.7	0.0155	5.8
PT-W-B	Time * education	−0.0234	5.1	−0.0280	5.5	−0.0299	5.3	−0.0061	1.6
LOS Variables (random variation)									
All	Time	−0.0098	3.7	−0.0102	4.0	−0.0222	4.1	−0.0405	7.5
	SD			−0.0006	0.6	0.0287	5.3	0.0764	17.0
All	Cost	−0.2044	8.0	−0.3156	8.7	−0.4369	9.4	−0.2956	13.2
	SD			0.1935	6.3	0.3415	8.6	0.4495	13.4
N observations		3,214		3,214		3,214		3,214	
N repetitions/individual		3,214		3,214		2,048		100	
Log likelihood (0)		−3,397.60		−3,397.60		−3,397.60		−3,397.60	
Final log likelihood		−2,321.35		−2,310.81		−2,269.68		−1,978.88	
Adjusted rho-squared		0.310		0.312		0.325		0.410	

[a]CD = car driver, CP = car passenger, PT = public transport, W = walking, and B = bike.

would be the best; hence, failing to account for heterogeneity and correlation effects produces erroneous results and wrong conclusions in future scenario policy analysis. However, if one believes the validation results, the effect would be the opposite.

CONCLUSIONS

In this paper issues related to model validation and policy analysis when observations are extracted from the multiday/multiweek travel survey have been explored. The problem is quite relevant because these data sets are characterized by a small number of respondents and repeated observations during a period that is usually 1 week long, but can be as long as 6 weeks. The topic treated is even more important if it is considered that researchers in activity-based modeling are trying to extend the usual 1-day framework to longer periods of time and that dynamic effects are now included in the model formulation to account for past history, habit, and state-dependency effects.

The analyses provided are based on simulated and real data. Two dimensions have been considered for model validation: the individual and the week (or subset of responses). Results from simulated data indicate that when the model is estimated on a subset of individuals, formulations accounting for heterogeneity and correlation effects (which correspond to the true model structure) are able to

TABLE 6 MobiDrive Validation: Modal Shares

Alternative	Observed	Predicted			
		Logit	ML-Cross	ML-Panel_d	ML-Panel_i
Sample: 100 Individuals and 3,214 Observations					
CD	375	343	339	332	285
CP	119	114	116	119	121
PT	59	85	90	100	120
W	125	163	158	153	178
B	196	170	170	170	170
D		128	129	138	232
2-norms		62.1	63	71	124
Sample: 126 Individuals and 3,319 Observations					
CD	287	270	261	255	236
CP	66	96	97	99	103
PT	124	137	142	148	142
W	119	110	111	108	122
B	154	135	138	139	146
D	—	88	100	114	118
2-norms	—	42.4	47.9	54.5	66.4

provide better forecasts, although not dramatic. Including heterogeneity and correlation effects does not improve forecasts when models are estimated on the full set of individuals, but on a subset of responses from each individual. These findings are not confirmed by results obtained estimating a mode choice model on a real 6-week travel diary. Models accounting for taste heterogeneity in time and cost, correlation over daily tours, and correlation over individual tours do not provide better forecasts with respect to simple multinomial logit; it is indicated instead that the correlation does not have an effect on the model's ability to reproduce real choices.

How individual choices change when level-of-service variables vary according to transportation policies has also been studied. Simulated cases reveal that ignoring heterogeneity and correlation effects affect the model's ability to recover the real modal shifts; in one case the model that accounts just for taste heterogeneity is already able to reproduce the real changes in probabilities. Modal shifts in the real case study differ substantially with the model formulations, but nothing can be said about the errors committed when heterogeneity and/or correlation effects are ignored. To conclude, models with better fit do not necessarily provide better forecasts, and ignoring taste heterogeneity can cause major bias in modal shifts calculation.

TABLE 7 MobiDrive Policy Analysis, 100 Individuals and 3,214 Observations

Alternative	Logit	ML-Cross	ML-Panel_d	ML-Panel_i
+30% Time Alternative 1 and Alternative 2				
CD	−0.011	−0.011	−0.017	−0.025
CP	−0.019	−0.019	−0.037	−0.050
PT	0.021	0.020	0.034	0.055
W	0.002	0.002	0.001	−0.002
B	0.007	0.007	0.011	0.011
+50% Cost Alternative 1				
CD	−0.056	−0.072	−0.081	−0.054
CP	0.045	0.066	0.085	0.058
PT	0.063	0.065	0.058	0.016
W	0.004	0.006	0.008	0.007
B	0.019	0.032	0.042	0.035

This analysis has been carried out with current but well-recognized modeling tools; more complex model specifications accounting for different dimensions of correlation [as in Cherchi et al. (*20*) and Hess and Rose (*30*)] could be adopted, although the computation time for model estimation is expected to rise with the temporal dimensions considered. The authors are considering extending this research work to other choice models in activity-based model systems: number of activities, activity participation, and time-of-day choice.

REFERENCES

1. Golob, T. F. Structural Equations Modeling of the Dynamics of Travel Choice Dynamics. In *Development in Dynamic and Activity-Based Approaches to Travel Analysis* (P. Jones, ed.), Gower Publishing Co., Brookfield, Vt., 1990, pp. 343–383.
2. Bradley, M. A Practical Comparison of Modeling Approaches for Panel Data. In *Panels for Transportation Planning: Methods and Applications* (T. F. Golob, R. Kitamura, and L. Long, eds.), Kluwer Academic Publishers, Norwell, Mass., 1997, pp. 281–304.
3. Simma, A., and K. W. Axhausen. Commitments and Modal Usage: Analysis of German and Dutch Panels. In *Transportation Research Record: Journal of the Transportation Board, No. 1854,* Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 22–31.
4. Chatterjee, K., and K. Ma. Behavioural Responses to a New Transport Option—A Dynamic Analysis Using a Six-Month Panel Survey. Presented at 11th International Conference on Travel Behaviour Research, Kyoto, Japan, 2006.
5. Thørgersen, J. Understanding Repetitive Travel Mode Choice in a Stable Context: A Panel Study Approach. *Transportation Research Part A,* Vol. 40, No. 8, 2006, pp. 621–638.
6. Srinivasan, K., and P. Bhargavi. Longer-Term Changes in Mode Choice Decisions in Chennai: A Comparison Between Cross-Sectional and Dynamic Models. *Transportation,* Vol. 34, No. 3, 2007, pp. 355–374.
7. Ramadurai, G., and K. K. Srinivasan. Dynamics and Variability in Within-Day Mode Choice Decisions: Role of State Dependence, Habit Persistence, and Unobserved Heterogeneity. In *Transportation Research Record: Journal of the Transportation Board, No. 1977,* Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 43–52.
8. Yañez, M. F., and J. de D. Ortúzar. A Panel Data Model to Forecast the Effect of a Radical Public Transportation Innovation. Presented at 4th International Symposium on TDM 2008, Vienna, Austria, 2008.
9. Axhausen, K. W., A. Zimmermann, S. Schönfelder, G. Rindsfüser, and T. Haupt. Observing the Rhythms of Daily Life: A Six-Week Travel Diary. *Transportation,* Vol. 29, No. 2, 2002, pp. 95–124.
10. Schlich, R., and K. W. Axhausen. Habitual Travel Behaviour: Evidence from a Six-Week Travel Diary. *Transportation,* Vol. 30, No. 1, 2003, pp. 13–36.
11. Stopher, P. R., E. Clifford, and M. A. Montes. Variability of Travel over Multiple Days: Analysis of Three Panel Waves. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2054,* Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 56–63.
12. Cirillo, C., and K. W. Axhausen. Evidence on the Distribution of Values of Travel Time Savings from a Six-Week Travel Diary. *Transportation Research Part A,* Vol. 40, No. 5, 2006, pp. 444–457.
13. Cherchi, E., and C. Cirillo. Variations in Responses and Mixed Logit Mode Choice Model Using Panel Data: Accounting for Systematic and

Random Preferences. Presented at 87th Annual Meeting of the Transportation Research Board, Washington, D.C., 2008.
14. Bhat, C. R., and S. Srinivasan. A Multidimensional Mixed Ordered-Response Model for Analyzing Weekend Activity Participation. *Transportation Research Part B,* Vol. 39, No. 3, 2005, pp. 255–278.
15. Habib, K. M. N., and E. J. Miller. Modelling Daily Activity Program Generation Considering Within-Day and Day-to-Day Dynamics in Activity-Travel Behaviour. *Transportation,* Vol. 35, No. 4, 2008, pp. 467–484.
16. Vanhulsel, M., D. Janssens, and G. Wets. Calibrating a New Reinforcement Learning Mechanism for Modeling Dynamic Activity-Travel Behavior and Key Events. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C., 2007.
17. Arentze, T., and H. J. P. Timmermans. A Dynamic Model for Generating Multi-Day, Multi-Person Activity Agendas: Approach and Illustration. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C., 2007.
18. Cirillo, C., and K. W. Axhausen. Dynamic Model of Activity-Type Choice and Scheduling. *Transportation,* Vol. 37, No. 1, 2010, pp. 15–38.
19. Yañez, M. F., E. Cherchi, B. G. Heydecker, and J. de D. Ortúzar. On the Treatment of the Repeated Observations in Panel Data: Efficiency of Mixed Logit Parameter Estimates. *Network and Spatial Economics,* 2010 (in press).
20. Cherchi, E., C. Cirillo, and J. de D. Ortúzar. A Mixed Logit Mode Choice Model on Panel Data: Accounting for Different Correlation over Time Periods. Presented at International Conference on Discrete Choice Modelling. Leeds, United Kingdom, March 2009.
21. Train, K. E. *Discrete Choice Methods with Simulation.* Cambridge University Press, United Kingdom, 2003.
22. Revelt, D., and K. Train. Mixed Logit with Repeated Choices. *Review of Economics and Statistics,* Vol. 80, No. 4, 1998, pp. 647–657.
23. Bastin, F., C. Cirillo, and P. L. Toint. Application of an Adaptive Monte-Carlo Algorithm for Mixed Logit Estimation. *Transportation Research Part B,* Vol. 40, No. 7, 2006, pp. 557–593.
24. Cherchi, E., and J. W. Polak. Assessing User Benefits with Discrete Choice Models: Implications of Specification Errors Under Random Taste Heterogeneity. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1926,* Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 61–69.
25. Munizaga, M. A., B. G. Heydecker, and J. de D. Ortúzar. Representation of Heteroscedasticity in Discrete Choice Models. *Transportation Research Part B,* Vol. 34, No. 3, 2000, pp. 219–240.
26. PTV AG, B. Fell, S. Schönfelder, and K. W. Axhausen. MobiDrive questionnaires, *Arbeitsberichte Verkehrs- und Raumplanung, 52.* Institut für Verkehrsplanung, Transporttechnik, Strassen- und Eisenbahnbau, ETH Zürich, Switzerland, 2000.
27. Schönfelder, S. *Urban Rhythms: Modelling the Rhythms of Individual Travel Behaviour.* Department Bau, Umwelt und Geomatik, ETH Zürich, Switzerland, 2006.
28. Bhat, C. R., and S. K. Singh. A Comprehensive Daily Activity-Travel Generation Model System for Workers. *Transportation Research A,* Vol. 34, No. 1, 2000, pp. 1–22.
29. Cherchi, E., and J. de D. Ortúzar. Can Mixed Logit Reveal the Actual Data Generating Process? Some Implications for Environmental Assessment. *Transportation Research D,* Vol. 15, No. 7, 2010, pp. 428–442.
30. Hess, S., and J. Rose. Allowing for Intra-Respondent Variations in Coefficients Estimated on Repeated Choice Data. *Transportation Research Part B,* Vol. 43, No. 6, 2009, pp. 708–719.

The Transportation Demand Forecasting Committee peer-reviewed this paper.

Defining Interalternative Error Structures for Joint Revealed Preference–Stated Preference Modeling

New Evidence

María Francisca Yáñez, Elisabetta Cherchi, and Juan de Dios Ortúzar

Joint model estimation with revealed (RP) and stated preference (SP) data has become popular in the past few years, and it is now considered common practice. Many theoretical issues related to estimation and prediction with joint RP-SP data are far from being fully explored. Given the ample diffusion of RP-SP modeling in practice, its misuse can have severe consequences on policy analysis and evaluation of transport investments; thus, it is crucial to continue research on this problem to try to give a theoretical justification to the many relevant issues that remain uncovered. One particularly interesting issue, which has not been well explored, is the effect of partial data enrichment on the correlation structure of alternatives (i.e., when different correlation structures are revealed in the RP and SP data sets). This problem, which is often found in practice, has no trivial solution and raises new interesting theoretical questions about estimation and prediction. In this paper, theoretical and practical implications of this problem are discussed and then empirical evidence is provided, from a real case, of the errors that may creep in when these models are not applied correctly. Finally some guidelines to help fill this important gap in the proper use of RP-SP data are provided.

Modern life depends heavily on the correct and efficient workings of the transport system. However, improving current transport services, or introducing new transport systems, requires large amounts of money. Thus, it is essential to count on consistent analysis tools to avoid unjustified investments. The past decade has been characterized by an intense and fruitful research effort on travel demand modeling, with the common objective to improve our capability to evaluate policy effects over users' choices. However, despite the practical relevance of joint revealed preference (RP) stated preference (SP) modeling, theoretical research on this issue has almost disappeared from the recent literature.

After the pioneering work of Ben-Akiva and Morikawa, the mixed RP-SP approach experienced a period of great popularity, in which several advances were made (*1*). The research produced in this first period involved fairly simplistic assumptions for the error structure, such as the simple multinomial logit (MNL) model or at most a nested logit (NL) structure for the RP alternatives and an MNL model for the SP data (*2–6*). Still by using this simple NL(RP)–MNL(SP) structure, some more recent works have explored the differences in preference structures among RP and SP data (*7*).

A major boost to the joint RP-SP modeling paradigm was given by the advent of the mixed logit model (*8*). In particular, mixed RP-SP models were estimated that incorporated the correlation due to repeated observations from the same individual and heteroscedasticity (*9*), random parameters under the full data enrichment approach (i.e., the RP and SP attributes have the same parameters) (*9–11*), and random parameters under the partial data enrichment approach (i.e., specific parameters for the RP and SP attributes) (*9, 12, 13*).

Another stream of research has referred to the reference dependence effect. The pioneer in dealing with state dependence and serial correlation for mixed RP-SP data is Morikawa (*14*). Then some studies suggested a practical approach that consisted of introducing a dummy inertia variable to indicate when the RP and SP sets shared the same choice (*3, 15*). Lately, Cantillo et al. proposed a more complex model, allowing estimating inertia as a function of the previous valuation of the alternatives, which also allowed for a consideration of serial correlation (*16*).

The majority of these works, however, focus only on model estimation. Forecasting is not an issue when RP and SP models share the same structure (*11, 17*). Problems arise instead when the RP and SP structures are not consistent because that hampers the whole procedure of passing from estimation to forecasting (*13*). It is useful to remember that for prediction, only the RP environment should be considered because it represents "real" behavior. Thus, even if a joint RP-SP model is built to obtain better estimates, for forecasting all information must be moved to the RP environment. Papers found in the literature basically discuss two problems: (*a*) how to transfer parameters associated with level-of-services variables (*5, 7, 9, 12*) and (*b*) how to transfer the alternative specific constants (ASC) (*12, 18, 19*).

Interestingly, although one of the major aims of using SP data is to test the introduction of new alternatives, only a few papers approach the problem of using RP-SP jointly estimated models to forecast the demand for new alternatives (*12, 18, 19*). Moreover, no paper was found that discussed the problem of using joint RP-SP models in prediction when the SP data included new alternatives that were also correlated. The authors believe that the origin of this void might be the dominant presence, at least in joint RP-SP modeling, of simplified SP designs based mainly on binary choices. Moreover, even when a higher number of alternatives are included

M. F. Yáñez and J. de D. Ortúzar, Department of Transport Engineering and Logistics, Pontificia Universidad Católica de Chile, Casilla 306, Cod. 105, Santiago 22, Chile. E. Cherchi, Research Center in Mobility Patterns, Department of Engineering, University of Cagliari, Piazza d'Armi 16, 09123 Cagliari, Italy. Corresponding author: M. F. Yáñez, mfyanezc@puc.cl.

Transportation Research Record: Journal of the Transportation Research Board, No. 2175, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 65–73.
DOI: 10.3141/2175-08

in the SP data, correlation among alternatives is not estimated (*20, 21*) or estimated but not considered in forecasting (*4, 9, 22*). Hensher and Rose report an interesting example of correlation among SP alternatives, but not in the context of joint RP-SP estimation (*23*).

Given this context, the aim of this paper is to discuss the theoretical problems involved in estimating and forecasting demand when a new group of correlated alternatives is introduced into the market. This is an interesting problem because (*a*) it raises new interesting theoretical issues about both estimation and prediction, (*b*) it is very likely to occur in practice, and (*c*) it may produce severe errors if not approached correctly.

The rest of the paper is organized as follows. Following is a brief review of the classical joint estimation for RP-SP data and a discussion of the problem of partial data enrichment in the correlation structure among alternatives. The next section provides a short description of the database used for the analysis and reports some empirical results. Finally, a summary of the main conclusions end the paper.

ESTIMATION AND FORECASTING WITH JOINT RP-SP DATA

As is well known, using mixed RP-SP data to estimate discrete choice models does not mean "simply join the data" because the scale factor in the indirect utility function must be considered (*2, 5*). The scale factor depends on the standard deviation of the error terms in the sample; hence it reflects the unobserved variability among individuals in the difference between the true and the estimated phenomenon. Because of the scale factor, two identical models estimated with different data sets may give different estimated parameters, even if the individual choice process is the same. In the case of RP-SP data sets, the typical random utility functions of the alternative j for the individual q, can be written as

$$\begin{aligned} U_{qj}^{RP} &= \beta X_{qj}^{RP} + \theta^{RP} Y_{qj} + \epsilon_{qj}^{RP} \qquad \epsilon_{qj}^{RP} \approx \left(o, \sigma_{RP}^2\right) \\ U_{qj}^{SP} &= \beta X_{qj}^{SP} + \theta^{SP} Z_{qj} + \epsilon_{qj}^{SP} \qquad \epsilon_{qj}^{SP} \approx \left(o, \sigma_{SP}^2\right) \end{aligned} \tag{1}$$

where

X^{RP} and X^{SP} = vectors of common attributes to both data sets (RP and SP);
β = corresponding vector of parameters;
Y and Z = vectors of attributes specific to each type of data, the parameters of which are, respectively, θ^{RP} and θ^{SP}; and
ϵ^{RP} and ϵ^{SP} = random terms associated with the RP and SP utilities, respectively, with variances σ_{RP}^2 and σ_{RP}^2, which will, in general, be different.

The index for SP pseudoindividuals is omitted in Equation 1.

The joint RP-SP estimation is a typical case of heteroscedasticity among subgroups of individuals (Q^{RP} and Q^{SP}). Hence, the general way to handle the problem consists of estimating different scale parameters for each data set (*24*). Because both variances cannot be identified, one needs to be normalized. However, estimating a heteroscedastic model normalized by the RP scale parameter (λ^{RP}) is equivalent to scaling the SP utility by a coefficient $\mu_{SP} = \lambda^{SP}/\lambda^{RP}$. This is precisely the first method suggested by Ben-Akiva and Morikawa to tackle joint estimation, where the coefficient (μ_{SP}) was justified by the need to have equal variances for the two subgroups of data ($\mu_{SP}^2 \cdot \sigma_{SP}^2 = \sigma_{RP}^2$) (*1*). As the error variance in an MNL model is $\sigma^2 = \pi^2/6\lambda^2$, the following equalities hold for the SP scale parameter in that case: $\mu_{SP} = \sigma^{RP}/\sigma^{SP} = \lambda^{SP}/\lambda^{RP}$. Hence, it is common to visualize the estimated parameters as either multiplied by the unknown scale parameter or divided by the unknown standard deviation. It will be seen later that the same equality does not apply when one goes beyond the MNL model.

Once accounting for the heteroscedasticity between the two data sets, either by scaling the SP utility by μ_{SP} or by formulating a heteroscedastic model (where λ^{SP} is estimated normalized by the RP factor scale), the log likelihood function for the joint estimation is given simply by the product of the individual probabilities of selecting the alternative actually chosen in the two data sets. The following two expressions are equivalent:

$$\begin{aligned} L &= \prod_{q \in Q^{RP}} P_{qj}^{g_{qj}} \left(\lambda^{RP} V_{qj}^{RP}\right) \cdot \prod_{q \in Q^{SP}} P_{qj}^{g_{qj}} \left(\mu_{SP} \left(\lambda^{RP} V_{qj}^{SP}\right)\right) \\ L &= \prod_{q \in Q^{RP}} P_{qj}^{g_{qj}} \left(\lambda^{RP} V_{qj}^{RP}\right) \cdot \prod_{q \in Q^{SP}} P_{qj}^{g_{qj}} \left(\lambda^{SP} V_{qj}^{SP}\right) \end{aligned} \tag{2}$$

where V_{qj}^{RP} is the representative utility for alternative j for individual q for RP observations and g_{qj} takes the value of 1 if individual q chooses alternative j; 0 otherwise. In both cases, the overall likelihood function is scaled by the unknown (inestimable) λ^{RP} scale factor. This means that in the joint RP-SP estimation all the estimated parameters (θ) are deflated by the common λ^{RP}.

Equation 2 is intentionally extremely generic because the joint estimation does not imply any restriction on the error structure across the RP and SP environments. In this paper the focus will be on the case of interalternative correlation in the RP or SP data sets or both; this is an important issue that has never been addressed. As will be discussed, the problem consists mainly in identifying the correct role of the scaling parameters to estimate a consistent model and then correctly applying the model in forecasting.

As is well known, correlation among subgroups of alternatives can be accounted for by using an error component (EC) structure or resorting to an NL model (*25, 26*) in the generalized extreme value (GEV) setting (*27*). The EC structure has the advantage, over the NL, of being more flexible and allowing in principle for intra-individual correlation; however, in practice intra-individual correlation induces correlation over all alternatives masking the correlation in the subgroup. The problem can be overcome by using more complex EC structures, but these suffer from theoretical identification problems (*10*). In the case of null intra-individual correlation, of course, the standard NL is preferable because it is far easier to estimate and its results to be interpreted.

As discussed in the introduction, although there are several studies on the estimation with both structures (EC and NL), neither examples of different interalternative correlation between RP and SP data nor a theoretical discussion on model scaling could be found, which are crucial to understand later how to use the models in prediction.

The EC structure consists of simply adding to each of the correlated alternatives a common term distributed with mean zero and variance to be estimated. In regard to estimated parameters, the variance of the EC structure is like any other parameter. Hence if the following EC structure [where ν_{ij} is distributed $N(0, 1)$, and it is equal for alternatives i and j] is included in the SP component (keeping the RP component as above),

$$\begin{aligned} U_{qj}^{RP} &= \beta X_{qj}^{RP} + \theta^{RP} Y_{qj} + \epsilon_{qj}^{RP} \qquad \epsilon_{qj}^{RP} \approx \left(o, \sigma_{RP}^2\right) \\ U_{qj}^{SP} &= \beta X_{qj}^{SP} + \theta^{SP} Z_{qj} + \epsilon_{qj}^{SP} + \nu_{qij}\sigma_{ij} \qquad \epsilon_{qj}^{SP} \approx \left(o, \sigma_{SP}^2\right) \end{aligned} \tag{3}$$

it is easy to see that its standard deviation (σ_{ij}) estimate will be deflated by the RP scale factor of the RP error terms (θ^{RP}).

The RP-SP scaling of an interalternative correlation in the GEV setting is less intuitive. Suppose that a set of alternatives can be grouped into K not overlapping nests (N_k) and that alternative j belongs to a particular nest N_k. Let λ_k also be the scale parameter associated with nest N_k and β the homogeneity parameter (i.e., scale factor at the root) of the GEV model. The GEV probability of choosing alternative j is equal to the product of the conditional probability of choosing nest N_k and the marginal probability of choosing it inside nest N_k:

$$P_{qj(j\in N_k)} = \frac{\exp(\lambda_k V_{qj})}{\sum_{j\in N_k}\exp(\lambda_k V_{qj})} \cdot \frac{\exp\frac{\beta}{\lambda_k}\left(\ln\sum_{j\in N_k}\exp(\lambda_k V_{qj})\right)}{\sum_{i=1,\ldots,k}\exp\frac{\beta}{\lambda_i}\left(\ln\sum_{j\in N_i}\exp(\lambda_i V_{qj})\right)} \tag{4}$$

In Equation 4 that part of the utility that depends only on variables that differ over nests but not over alternatives within each nest was omitted. The omission simplifies the notation without affecting the discussion. It is also important to remember that Equation 4 is equivalent to the joint probability that comes directly from the GEV structure [see Train for a demonstration (*8*)]. As is well known, the NL model has more than one scale parameter. The parameter chosen to be normalized does not affect the estimation. However, as noted by Carrasco and Ortúzar, the normalization has an impact on interpretation; only under the "upper normalization" can the NL be compared directly with an MNL (*26*). Moreover, under the upper normalization Equation 4 is also equivalent to the utility maximizing nested logit (*27*). Also, a "lower normalization" (i.e., one of the $\lambda_k = 1$) is consistent with microeconomic theory only when all the parameters λ_k associated with the stochastic errors within each nest k (and their respective structural parameters) have the same value. This is relevant particularly in the joint RP-SP estimation, and it may also be one of the reasons that the literature reports mainly very simple structures (i.e., MNL structures or NL structures with at most one nest). However, the lower normalization may have an advantage for practitioners because, in these common simple structures, scaling the estimated parameters is not required. In the following subsections the consistency problem of joint RP-SP models in estimation and the implication in forecasting are discussed.

Consistency in RP-SP Model Estimation

First, it is useful to report the expression of the joint RP-SP likelihood for the most general case (i.e., with NL models in the RP and SP settings) without any assumption about the normalization. By applying Equation 4 to both data sets, the likelihood would be as follows:

$$L = \prod_{q\in Q^{RP}} \frac{\exp(\lambda_k^{RP} V_{qj}^{RP})}{\sum_{j\in N_k^{RP}}\exp(\lambda_k^{RP} V_{qj}^{RP})} \cdot \frac{\exp\frac{\beta^{RP}}{\lambda_k^{RP}}\left(\ln\sum_{j\in N_k^{RP}}\exp(\lambda_k^{RP} V_{qj}^{RP})\right)}{\sum_{i=1,\ldots,K^{RP}}\exp\frac{\beta^{RP}}{\lambda_i^{RP}}\left(\ln\sum_{j\in N_i^{RP}}\exp(\lambda_i^{RP} V_{qj}^{RP})\right)}$$
$$\cdot \prod_{q\in Q^{SP}} \frac{\exp(\lambda_k^{SP} V_{qj}^{SP})}{\sum_{j\in N_k^{SP}}\exp(\lambda_k^{SP} V_{qj}^{SP})} \cdot \frac{\exp\frac{\beta^{SP}}{\lambda_k^{SP}}\left(\ln\sum_{j\in N_k^{SP}}\exp(\lambda_k^{SP} V_{qj}^{RP})\right)}{\sum_{i=1,\ldots,K^{SP}}\exp\frac{\beta^{SP}}{\lambda_i^{SP}}\left(\ln\sum_{j\in N_i^{SP}}\exp(\lambda_i^{SP} V_{qj}^{RP})\right)} \tag{5}$$

RP-NL(equal scale factors)–SP-MNL

This is by far the most used structure, although not the simplest (as that is the case in which both structures are MNL, Equation 2). The most common case is the one in which the RP data present only one subgroup of correlated alternatives, the SP data include only independent alternatives, and both data sets are homoscedastic (which means $\lambda_k^{RP} = \lambda^{RP}$; $\forall k$, $\lambda_k^{SP} = \lambda^{SP}$; $\forall k$ and, more important, $\beta^{SP} = \lambda^{SP}$). In that case, Equation 5 reduces to

$$L = \prod_{q\in Q^{SP}} \frac{\exp(\lambda^{RP} V_{qj}^{RP})}{\sum_{j\in N_k}\exp(\lambda^{RP} V_{qj}^{RP})} \cdot \frac{\exp\frac{\beta^{RP}}{\lambda^{RP}}\left(\ln\sum_{j\in N_k}\exp(\lambda^{RP} V_{qj}^{RP})\right)}{\sum_{i=1,\ldots,K}\exp\frac{\beta^{RP}}{\lambda^{RP}}\left(\ln\sum_{j\in N_i}\exp(\lambda^{RP} V_{qj}^{RP})\right)}$$
$$\cdot \prod_{q\in Q^{RP}} \frac{\exp(\lambda^{SP} V_{qj}^{SP})}{\sum_{j\in N_k}\exp(\lambda^{SP} V_{qj}^{SP})} \tag{6}$$

All cases above refer to the alternatives effectively chosen at each environment; also, although only one scale factor (λ^{RP}) is assumed in the RP data set, because the SP data have different scale from the RP data, the SP scale factor (λ^{SP}) needs to be explicitly estimated to account for heteroscedasticity. But, as discussed previously, the SP utilities are scaled by one or other RP scale parameter depending on the normalization chosen. In particular, $\hat{\lambda}^{SP}\hat{V}_{qj}^{SP} = \lambda^{SP}/\beta^{RP}(\beta^{RP} V_{qj}^{SP})$ will be estimated (the hat indicates the estimated parameter) under the upper normalization (where $\lambda^{RP} = 1$; hence λ^{RP} is the unknown RP scale and $\hat{\lambda}^{SP}$ is the estimated SP scale), and $\hat{\lambda}^{SP}\hat{V}_{qj}^{SP} = \lambda^{SP}/\lambda^{RP}(\lambda^{RP} V_{qj}^{SP})$ will be estimated under the lower normalization.

In this latter case, the ratio $\lambda^{SP}/\lambda^{RP}$ represents the $\hat{\mu}_{SP}$ parameter that is multiplied by the SP utilities to make the whole RP-SP structure homoscedastic under Ben-Akiva and Morikawa's framework (*1*). However, because the NL-RP variance is $\sigma^2_{\beta RP}$ (with $\sigma^2_{\beta RP} = \sigma^2_{RP} + \sigma^2_{\lambda RP}$, and σ^2_{iRP} has a logistic distribution) and the MNL-SP variance is ($\sigma^2_{\beta SP} = \sigma^2_{\lambda SP}$), the condition for variance equality becomes $\mu^2_{SP}\sigma^2_{\lambda SP} = \sigma^2_{\beta RP}$. Hence, only the upper normalization is consistent with the condition of equal variance. To obtain the same consistency under the lower normalization, the SP utilities need to be multiplied by two parameters: the SP scale parameter ($\hat{\mu}_{SP} = \lambda^{SP}/\lambda^{RP}$) and the inverse of the RP-NL structural parameter (λ^{RP}/β^{RP}). Note also that this is actually the "trick" used under nonnormalized NL (NNNL) to obtain a consistent structure (*28*). The reason is that the NNNL uses a lower normalization.

RP-NL(different scale factors)–SP-MNL

Although it is not uncommon to find NL models with different scale factors for each nest using a single data set, only one example of estimating these structures jointly with SP data has been found (*22*). In this case the advantage of the upper normalization is even clearer because the SP data set needs only to be scaled by the ratio λ^{SP}/β^{RP}. Under the lower normalization instead, suppose that one normalizes for $\lambda_k^{RP} = 1$ for all RP alternatives that belong to any nest $i \neq k$, the following $K^{RP} - 1$ parameters would need to be estimated: $(\lambda_i^{RP}/\lambda_k^{RP})$ [these are equivalent to the ratio (ϕ_r/ϕ_j') used by Carrasco and Ortúzar (*26*), Equation 23]. More than that, $(\beta^{RP}/\lambda_k^{RP})$ also needs to be computed plus the SP scale parameter $\lambda^{SP}/\lambda_k^{RP}$.

RP-MNL–SP-NL(equal scale factors)

This is a case that has never been discussed in the literature. It can typically occur when, for example, in the current situation there is only one public transport mode (say, bus) and the SP data are gathered to test the introduction of a new public mode (for example, tram) that will probably be correlated with the current bus system. In this case the RP alternatives are independent; hence only the RP upper normalization is possible because from Equation 4 this is $\beta^{RP} = \lambda^{RP} = \lambda_k^{RP}, \forall k$. Hence the two NL-SP parameters (β^{SP}, λ^{SP}) will be estimated scaled by the unknown λ^{RP} (or equivalently λ^{RP}) parameter.

$$L = \prod_{q \in Q^{SP}} \frac{\exp\left(\beta^{RP} V_{qj}^{RP}\right)}{\sum_{j \in N_k^{RP}} \exp\left(\beta^{RP} V_{qj}^{RP}\right)} \cdot \prod_{q \in Q^{RP}} \frac{\exp\left(\lambda^{SP} V_{qj}^{SP}\right)}{\sum_{j \in N_k^{SP}} \exp\left(\lambda^{SP} V_{qj}^{SP}\right)} \cdot \frac{\exp \frac{\beta^{SP}}{\lambda^{SP}} \left(\ln \sum_{j \in N_i^{SP}} \exp\left(\lambda^{SP} V_{qj}^{RP}\right) \right)}{\sum_{i=1,\ldots,K^{SP}} \exp \frac{\beta^{SP}}{\lambda^{SP}} \left(\ln \sum_{j \in N_i^{SP}} \exp\left(\lambda^{SP} V_{qj}^{RP}\right) \right)} \tag{7}$$

One can interpret the estimated parameter $\hat{\lambda}^{SP} = \lambda^{SP}/\beta^{RP}$ as the scale between RP and SP (what has so far been called $\hat{\mu}_{SP}$) and $\hat{\beta}^{SP} = \beta^{SP}/\beta^{RP}$ as the structural parameter among SP alternatives. This corresponds to a lower normalization into the SP data set. Alternatively, one can think of $\hat{\beta}^{SP} = \beta^{SP}/\beta^{RP}$ as the RP-SP scale and $\hat{\lambda}^{SP} = \lambda^{SP}/\beta^{RP}$ as the structural parameter under an upper normalization. The extension to more than one scale factor in the SP context is trivial, but of course the number of parameters to be estimated increases, being equal to the number of nests plus one: $\hat{\beta}^{SP} = \beta^{SP}/\beta^{RP}$ and $\hat{\lambda}_i^{SP} = \lambda_i^{SP}/\beta^{RP}$ for $i = 1, \ldots, K^{SP}$.

RP-NL–SP-NL(different scale factors)

This is the most general and complicated case, although again not rare to be found in practice. In fact, this is the case in which there are currently two transport modes, which might be correlated, and with the SP data it is desirable to test the inclusion of another one or two new modes, again probably correlated between them. An example, in an urban context, would be the introduction of a full new public transport system, such as a renewed bus service and a new underground. In a car-orientated city, it is likely that there will be correlation between the existent car driver and car passenger modes, as well as between the two new public modes. Analogously, in an extraurban context, an example can be the introduction of two new train services, for example, local and express train, between a pair origin–destination currently served only by two bus services, local and express buses.

The likelihood in this case is expressed by Equation 5. Once more, the upper normalization appears to be the most convenient because the joint RP-SP estimation is similar to the case discussed above in Point 3 [RP-MNL/SP-NL(equal scale factors)], but the number of parameters to estimate increases significantly, as they are $\hat{\lambda}_i^{RP} = \lambda_i^{RP}/\beta^{RP}$ for $i = 1, \ldots, K^{RP}$, $\hat{\beta}^{SP} = \beta^{SP}/\beta^{RP}$ and $\hat{\lambda}_i^{SP} = \lambda_i^{SP}/\beta^{RP}$ for $i = 1, \ldots, K^{SP}$.

Under the lower normalization instead, again assume one normalizes $\lambda_k^{RP} = 1$, the following parameters must be estimated:

- $\hat{\beta}^{RP} = \beta^{RP}/\lambda_k^{RP}$ in the RP environment,
- $K^{RP} - 1$ parameters $\hat{\lambda}_i^{RP} = \lambda_i^{RP}/\lambda_k^{RP}$ for all RP alternatives that belong to any nest $i \neq k$,
- $\hat{\beta}^{SP} = \beta^{SP}/\lambda_k^{RP}$ in the SP environment, and
- $K^{SP} - 1$ parameters $\hat{\lambda}_i^{SP} = \lambda_i^{SP}/\lambda_k^{RP}$ for all SP alternatives that belong to any nest $i \neq k$.

Rather than the number of parameters, in this case the problem attains the behavioral interpretation of such structure especially for those alternatives (or groups of correlated alternatives) that are common to the RP and SP data sets. Assuming different structural parameters in the RP and SP data sets means that the unobserved components of the utilities that cause some alternatives to be perceived as more similar than others are not the same in both cases. Although this can be justified when the systematic utilities are specified differently, it should not occur when alternatives have the same specification.

Forecasting with RP-SP Models

To use a joint RP-SP model estimated for prediction purposes one must take into account that real behavior belongs to the RP context. Therefore, when the model is used in forecasting, every parameter must be in the RP scale (*7*). When the joint RP-SP model is RP scaled, all parameters (i.e., RP-SP generic or specific to SP) associated with variables are already estimated in the RP environment; hence they can be used directly in prediction without further scaling. However, as discussed in the literature, this is not the case for the new-alternative ASC included only in SP (*12, 18, 19*). As the ASC tend to reproduce the market shares in the sample, they have to be rescaled when moved into the RP domain (*7, 19*).

Now, in the case of partial data enrichment in the interalternative correlation structure (i.e., the structural parameters are estimated only in the SP environment or in both the RP and SP settings but with different values), the following questions arise: (*a*) is it correct to move the whole (or part of the) SP correlation structure into the RP environment? and (*b*) can the structural parameters be moved free of scale or not? In regard to the first question, the following cases can occur:

1. The joint RP-SP structure assumes the same structural parameters for both environments. This is the simplest case, because there is no problem in using the structural parameters directly in forecasting. However, this case demands the same set of alternatives for both environments.
2. The joint RP-SP structure assumes the same structural parameters for both environments except for alternatives present only in SP (usually new alternatives). In this case one does not have an option other than moving the whole correlated (or uncorrelated) structure of SP alternatives into the RP domain to forecast.
3. The joint RP-SP structure allows different structural parameters for the two environments. This is the most general and most complicated case. First, the general advice would be to use the structure estimated with the data set (RP or SP) one has more faith in (*7*). However, passing judgment about the interalternative structural parameters in the GEV setting is not straightforward because these parameters are associated with the expected maximum utility (EMU) of the nests. Thus, to be consistent, the structural parameters should be associated with utilities measured in the same environment. In other words, for prediction all parameters (i.e., either structural or related to attributes) included in the joint RP-SP model should come from the same environment. This means that if one has more faith in the SP data, one should move the SP structural parameter and the utilities associated with the alternatives in the nest.

The authors' suggestion for existing alternatives is that, unless one has clear and specific reasons for not doing so (e.g., the goodness-of-fit statistics clearly reject it), it is preferable to estimate a generic correlation structure between RP and SP data. Moreover, for extreme cases in which interalternative correlation is present only in one environment (either RP or SP), that is, in the case of partial data enrichment in the interalternatives correlation, testing the following two structures is suggested: (*a*) constrain the RP alternatives to be correlated (as the SP ones) and (*b*) constrain the SP alternatives not to be correlated (as the RP ones).

As for the second question, when a group of correlated alternatives is estimated only in the SP environment, the SP utilities and the SP structural parameter need to be moved into the RP environment for prediction. Following the discussion in Cherchi and Ortúzar, the SP structural parameter should be moved without scaling it because it was estimated scaled by the RP scale parameter and because, different from the ASC that are not associated with any attributes, the SP structural parameter is associated with the EMU term (*7*).

EMPIRICAL RESULTS

Data Set

The data used in this paper were collected as part of a large project to study the introduction of a new high-speed (350 km/h) rail line between three large cities. The direct area of influence has a population of about 36 million inhabitants and in economic terms is one of the most important of the region. For the application of this research, only the data relative to one origin–destination pair were used because they represent a very interesting case study.

The two cities are located approximately 450 km from each other, and there are three modes available to travel between them: car, bus, and airplane. However, taking into account the clear competitiveness among the different services for the same mode, it can be said that there are actually two groups of three alternatives each: three bus alternatives (conventional bus, executive bus, sleeper bus) and three airplane alternatives (to represent three pairs of airports available).

The current transport market is dominated by an air shuttle service. Consequently, the local airports suffer from congestion. However, because the demand in the corridor is extremely high (some 120,000 daily trips), the slower modes (bus and car) are also affected by congestion, especially on approach roads that connect the city to the main arterial roadways. To have an idea of the problem, note that the journey by car and bus takes about 5 h whereas the air shuttle takes about 45 min between taxiing, taking off, and cruising, plus 50 min between arrival and boarding (for Internet/automatic check-in).

Demand forecasting for the new high-speed train was early identified as a key challenge in the project. Following the recommended approach, three types of surveys were set up: (*a*) focus groups to achieve a deeper understanding of the phenomenon, (*b*) an RP survey to obtain a good picture of the current situation, and (*c*) a stated choice experiment to evaluate the introduction of the new alternatives.

The RP survey considered some 2,050 individuals (1.7% of existing demand) interviewed at highway toll stations, bus terminals, and airports. Apart from standard data about trip conditions and socioeconomic characteristics, a question about car availability was included in the RP survey; the public modes were considered available to everyone. From the original sample some 1,760 individuals answered the SP survey satisfactorily. The SP experiment was designed to include all existing modes plus the new train services (economy and first class). However, to simplify the setting, each respondent was presented with only four of these modes. These were randomly selected but with the constraint that the current mode and the speed rail modes were present at every scenario. Moreover, each individual was presented with only four scenarios, blocking out the more than 48 choice tasks of the experimental design. Interestingly, in spite of the strong similarity between the two new rail modes (they differed only on the onboard service), the SP survey presented them as competing alternatives, as this was what the planners were specifically interested in. For this reason they were also considered as two separate but correlated alternatives.

A final RP-SP sample of roughly 9,100 observations was available for estimation purposes, of which 59.6% were trips to work. Different models were estimated according to the main purpose of the trip: to work and not to work. In this paper only the results referring to the first case will be presented.

Empirical Analysis

As good practice dictates, before the RP-SP joint models were estimated, separate models were estimated for each data set to obtain the best model specification (*5*). In particular, by using each data set alone the following effects were tested: linear and nonlinear in the attributes utility functions (interactions between alternative attributes and square terms), taste heterogeneity (deterministic, via interaction between level-of-service and socioeconomic characteristics and purely random), and for the SP data set, the existence of correlation among the observations from the same individual were also tested.

As for the utility specification, in both the RP and SP data sets, it was found that only the linear effects were significant, with the exception of cost, which showed significant systematic heterogeneity as a result of the income level of the respondent. In fact, it was found that a variable cost/income performed much better than the linear cost attribute, in both the RP and the SP sets. It was also found that the cost parameter was significantly different across alternatives. It is important to mention that neither data set showed random taste heterogeneity or preferences for specific alternatives. It is even more important to highlight that the SP data set did not show significant correlation among answers from the same individual; the effect of correlation was tested by using Biogeme in two ways: with random parameters and including an error component (*29*).

Finally, the results from the best RP and SP models estimated alone were compared to identify candidate RP-SP generic attributes for the joint estimation. It was found that all common attributes differed only by the model scale, while significantly different ASC were estimated for all RP and SP alternatives. A pity, as in forecasting it is preferable to have generic ASC for alternatives that are common to both sets (*19*).

However, the most interesting part of this joint RP-SP model refers to the estimation of the interalternative correlation structure, which of course imposes the challenge of testing complex joint NL structures as discussed previously.

In particular, different interalternative correlation structures for the RP and SP data were found. The SP data presented a clear and strong correlation between the two train services and among the three bus alternatives (see Figure 1); whereas the interalternative correlation among the RP bus alternatives was not significant (far lower than the 95% confidence level).

Although from the estimation point of view having different correlation structures in RP and SP is not an issue, using such structures in forecasting might be quite problematic. Now, because prediction should be the final aim of most studies, following the discussion in the

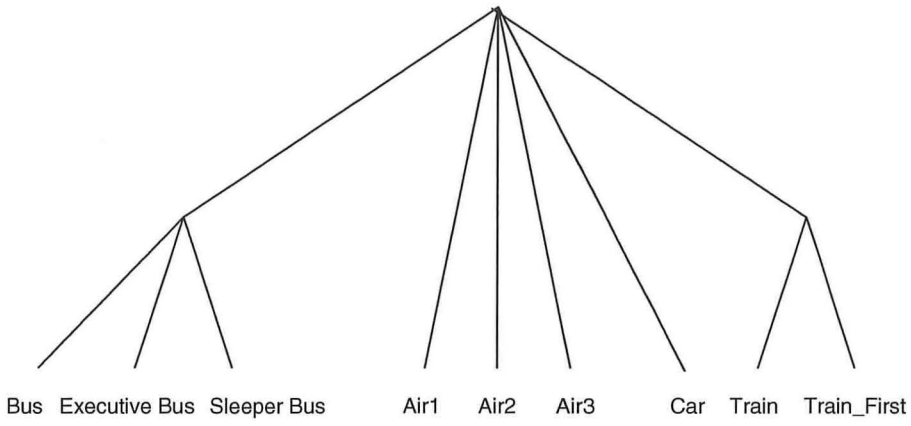


FIGURE 1 Nested model structure for SP data.

section on estimation and forecasting with joint RP-SP data, the correlation structure of both environments was constrained to establish a unique and consistent structure. This gave rise to three possible correlation structures for the joint RP-SP model: (*a*) all alternatives independent (Model MNL), (*b*) all alternatives independent except the two new alternatives in SP (Model MNL-NL Train), and (*c*) the bus alternatives correlated in both data sets with the same structural parameter and the two new alternatives correlated in SP (Model NL). The MNL model was estimated as the baseline for the nested specifications.

Table 1 presents the models tested in prediction; the parameters are already scaled as follows: according to Equation 5 and looking only at the RP side, λ_{bus} and λ_{train} correspond to $\beta^{RP}/\lambda^{RP}_{bus}$ and $\beta^{RP}/\lambda^{RP}_{train}$, respectively, whereas V^{RP}_{kj} can be calculated directly without scaling it again by λ^{RP}_k. Results from model estimation suggest that the structure revealed by the SP data is more appropriate. In fact, model performance increased when it is assumed that the bus alternatives in the RP set had the same correlation as that in the SP set. The likelihood ratio (LR) test allowed one to conclude that the best joint interalternative correlation structure (NL) accommodated correlation between the two new train modes with the same structural parameter for the RP and SP bus alternatives ($\phi^{SP}_{bus} = \phi^{RP}_{bus}$), whereas the MNL was the less appropriate model (*30*). In fact, the LR between the MNL and MNL-NL Train (LR = 46.2) is much larger than the critical value, $\chi^2_{95\%,1} = 3.84$.

In line with the LR test, the structural parameter comes closer to one as the interalternative correlation is better specified. It shows that part of the interalternative correlation is being masked by the scale parameter (for the deficient interalternative correlation specification). In addition, and to measure the "cost" of the constraint ($\phi^{SP}_{bus} = \phi^{RP}_{bus}$) allowing calibration of a consistent structure for prediction, the same correlation structure assuming different structural parameters for the bus alternatives ($\phi^{SP}_{bus} \neq \phi^{RP}_{bus}$) was also tested, but results were inconsistent because the structural parameter for the buses in the RP set was greater than one. Moreover, the likelihood of the inconsistent NL model was just slightly better (–5091.9). As to the issue of moving all information into the RP environment,

- Because different ASC were calibrated for the RP and SP contexts, the step of adjusting them to reproduce the current market shares was skipped;
- The ASC for the SP new alternatives were scaled; and
- The attribute parameters and the structural parameter were used directly (without scaling).

In regard to the relationship between travel time and access time, it can be seen that for the best model (NL) the travel time is (correctly, in the sense of what one normally expects) valued as 0.5 the access time for airplane and bus, and 0.6 for train. On the contrary, for the other two models (MNL and MNL-NL Train), the value of travel time is larger than the value of access time. Moreover, the simplest model (MNL) seems to strongly overestimate the value of travel time; it is six times the value of access time.

To evaluate the effect of establishing interalternative correlation on demand predictions, the variation in aggregate market shares for various simple policy measures was calculated. The response to a change in prediction was calculated (see Table 2) as the percent change in the aggregate share of mode *j* over the initial (do-nothing) situation:

$$\Delta P_j = \frac{P_j - P_j^0}{P_j^0} \tag{8}$$

where P_j and P_j^0 are the aggregate probabilities of choosing mode *j* before and after introducing the measure, both computed by sample enumeration (*30*).

As expected, all models predict a decrease in the market shares of the existing modes following a reduction in the train fares. However, there are important differences in the magnitude of the changes; in particular, the policy impacts are obviously greater for the MNL model. Table 2 shows that for a 50% reduction in train fares, if one erroneously assumes no interalternative correlation (MNL), the estimated percent change in the aggregate train share (ΔP_j) is 50% larger than if both train alternatives are assumed to be correlated. However, it is interesting that the differences between the two nested models are not large. This result could be due to the fact that the initial bus market share is low (17% for all three bus options) and its fare much lower than that of the competing modes. But, a similar phenomenon has been found in totally different contexts (*13*).

When there are either two (or more) correlated new alternatives, or a new alternative correlated with any existing one, the issue of

TABLE 1 Model Parameters for Prediction

Attribute	Models		
	MNL	MNL-NL Train	NL
Parameter			
Car	1.605	1.588	0.440
Air 1	−0.764	2.037	1.576
Air 2	−0.508	2.293	1.834
Air 3	−4.576	−1.776	−2.252
Sleeper bus	0.860	0.844	0.895
Executive bus	−1.201	−1.257	−1.094
Train	0.254	6.508	5.124
Train 1st class	0.363	6.632	5.216
Travel time car (min)	−0.012	−0.003	−0.001
Travel time air (min)	−0.012	−0.003	−0.001
Travel time bus (min)	−0.012	−0.003	−0.006
Travel time train (min)	−0.012	−0.008	−0.003
Access time air (min)	−0.002	−0.002	−0.002
Access time bus (min)	−0.002	−0.002	−0.013
Access time train (min)	−0.002	−0.005	−0.005
Cost/income air (US$)	−32.460	−7.163	−6.997
Cost/income bus (US$)	−29.327	−2.938	−19.789
Cost/income car (US$)	−77.867	−10.303	−10.010
Cost/income train economy (US$)	−51.640	−24.853	−23.521
Cost/income train first (US$)	−49.973	−24.853	−17.104
Delay air (min)	−0.013	−0.002	−0.0003
Delay bus (min)	−0.013	−0.002	−0.0022
Delay train (min)	−0.013	−0.002	−0.0009
ϕ_{Bus}	1.000	1.000	0.154
ϕ_{Train}	1.000	0.370	0.371
Scale parameter	0.136	0.5916	0.627
Final log likelihood	−5,121.01	−5,116.72	−5,093.62
Log likelihood at zero	−6,882.67	−6,712.7762	−6,712.78
Value of Time for Average Income (US$/h)			
Travel by car	19.81	37.43	12.84
Travel by air	70.41	79.76	27.22
Travel by bus	35.53	88.67	26.33
Travel by train economy	30.58	42.35	16.78
Travel by train first	31.60	42.35	23.08
Access air	11.73	53.18	54.44
Access bus	5.92	59.12	57.05
Access train economy	5.10	25.86	27.33
Access train first	5.27	26.47	38.46

considering a consistent interalternative correlation structure for prediction seems to be especially important because if the analyst wrongly assumes that the new alternatives are independent (MNL), their demand may be overestimated.

CONCLUSIONS

In this paper an analysis has been done of the problem of establishing an appropriate correlation structure for prediction when RP-SP data are used and the models in both environments present different correlation structures. From a theoretical point of view, the joint estimation of these structures poses a problem in defining an appropriate scale. In fact, when correlation among alternatives is present, several extraparameters need to be estimated to obtain a consistent joint model. Although lower and upper normalization are in principle equivalent, the theoretical discussion in this paper suggests that the upper normalization is more intuitive. In this sense the analyses add new evidence to support the results of Carrasco and Ortúzar related to NL estimation with only one type of data (*26*).

Although any interalternative structure between RP and SP can in principle be estimated, the theoretical analyses suggest that, wherever

TABLE 2 Effects on Prediction of Including Interalternative Correlation

	Attribute	Fare by Train			Fare by Airplane			Fare by Bus		
Model	% Change	–50	–25	–10	–50	–25	–10	–50	–25	–10
MNL	Car	–0.460	–0.230	–0.090	–0.390	–0.180	–0.060	–0.077	–0.039	–0.016
	Air	–0.410	–0.210	–0.080	1.050	0.460	0.170	–0.048	–0.024	–0.009
	Bus	–0.490	–0.230	–0.080	–0.400	–0.160	–0.050	0.157	0.078	0.031
	Train	0.630	0.310	0.120	–0.360	–0.170	–0.060	–0.057	–0.028	–0.011
MNL-NL Train	Car	–0.103	–0.051	–0.020	–0.139	–0.066	–0.026	–0.007	–0.003	–0.001
	Air	–0.081	–0.040	–0.016	0.200	0.093	0.036	–0.002	–0.001	–0.001
	Bus	–0.214	–0.108	–0.042	–0.215	–0.088	–0.031	0.048	0.024	0.010
	Train	0.139	0.070	0.027	–0.138	–0.066	–0.026	–0.005	–0.003	–0.001
NL	Car	–0.100	–0.050	–0.020	–0.140	–0.060	–0.030	–0.007	–0.003	–0.001
	Air	–0.080	–0.040	–0.010	0.190	0.090	0.030	–0.003	–0.001	–0.001
	Bus	–0.200	–0.100	–0.040	–0.210	–0.090	–0.030	0.050	0.025	0.010
	Train	0.130	0.070	0.030	–0.140	–0.070	–0.030	–0.006	–0.003	–0.001

possible, having a generic correlation structure for existing alternatives in both data sets is recommended. Otherwise, the model might not be consistent in prediction. Finally, if some interalternative correlation structure is present only in the SP data context (for new alternatives), it has to be moved to the RP environment without scaling the structural parameter for prediction.

It has been shown that models that simply follow the correlation structure detected for the RP data, without considering what the SP data might reveal in this sense, may overestimate the potential market shares of new alternatives. In other words, these results suggest that the preliminary step (estimation of pure RP and SP models) should be less restrictive to allow that the SP data may not only help in improving the specification of the representative utility, as in normal practice, but also in defining the most appropriate correlation structure.

These results may be also relevant for panel data modeling when the quality of data varies over waves. Thus, if one is in the presence of stable choice environments, one could assume a unique and consistent interalternative correlation structure for forecasting purposes.

ACKNOWLEDGMENTS

The authors thank the Millennium Institute on Complex Engineering Systems for funding. The paper was written while the first author was a visiting PhD student at the University of Cagliari and at the German Institute for Economic Research, financed by the Chilean Council for Scientific and Technological Research and the German Academic Exchange Service, and the last author was a visiting professor at the Centre for Transport Studies, University College London, under a research grant from the United Kingdom Science and Engineering Research Council.

REFERENCES

1. Ben-Akiva, M., and T. Morikawa. Estimation of Travel Demand Models from Multiple Data Sources. Presented at 11th International Symposium on Transportation and Traffic Theory, Yokohama, Japan, 1990.
2. Hensher, D. A. Stated Preference Analysis of Travel Choice: The State of Practice. *Transportation,* Vol. 21, No. 2, 1994, pp. 107–133.
3. Bradley, M., and A. Daly. Estimation of Logit Choice Models Using Mixed Stated Preference and Revealed Preference Information. In *Understanding Travel Behaviour in an Era of Change* (P. R. Stopher and M. Lee-Gosselin, eds.), Pergamon, Oxford, United Kingdom, 1997.
4. Ortúzar, J. de D., and A. Iacobelli. Mixed Modelling of Interurban Trips by Coach and Train. *Transportation Research Part A,* Vol. 32, No. 5, 1998, pp. 345–357.
5. Louviere, J. J., D. A. Hensher, and J. D. Swait. *Stated Choice Methods: Analysis and Application.* Cambridge University Press, United Kingdom, 2000.
6. Cherchi, E., and J. de D. Ortúzar. Mixed RP/SP Models Incorporating Interaction Effects. *Transportation,* Vol. 29, No. 4, 2002, pp. 371–395.
7. Cherchi, E., and J. de Dios Ortúzar. Use of Mixed Revealed-Preference and Stated-Preference Models with Nonlinear Effects in Forecasting. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1977,* Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 27–34.
8. Train, K. E. *Discrete Choice Methods with Simulation.* Cambridge University Press, Cambridge, United Kingdom, 2003.
9. Brownstone, D., D. S. Bunch, and K. Train. Joint Mixed Logit Models of Stated and Revealed Preferences for Alternative-Fuel Vehicles. *Transportation Research Part B,* Vol. 34, No. 5, 2000, pp. 315–338.
10. Walker, J. *Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures and Latent Variables.* PhD dissertation. Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Mass., 2000.
11. Bhat, C. R., and S. Castelar. A Unified Mixed Logit Framework for Modeling Revealed and Stated Preferences: Formulation and Application to Congestion Pricing Analysis in the San Francisco Bay Area. *Transportation Research Part B,* Vol. 36, No. 7, 2002, pp. 577–669.
12. Bhat, C., and R. Sardesai. The Impact of Stop-Making and Travel Time Reliability on Commute Mode Choice. *Transportation Research Part B,* Vol. 40, No. 9, 2006, pp. 709–730.
13. Cherchi, E., and J. de D. Ortúzar. On the Use of Mixed RP/SP Models in Prediction: Accounting for Random Taste Heterogeneity. *Transportation Science* (in press), 2010.
14. Morikawa, T. Correcting State Dependence and Serial Correlation in the RP/SP Combined Estimation Method. *Transportation,* Vol. 21, No. 2, 1994, pp. 153–165.
15. Brownstone, D., D. S. Bunch, T. Golob, and W. Ren. A Transactions Choice Model for Forecasting Demand for Alternative-Fuel Vehicles. In *Research in Transportation Economics* (S. McMullen, ed.), JAI Press, Inc., Greenwich, Conn., 1996.
16. Cantillo, V., J. de D. Ortúzar, and H. C. W. L. Williams. Modelling Discrete Choices in the Presence of Inertia and Serial Correlation. *Transportation Science,* Vol. 41, No. 2, 2007, pp. 195–205.

17. Espino, R., J. de D. Ortúzar, and C. Román. Understanding Suburban Travel Demand: Flexible Modelling with Revealed and Stated Choice Data. *Transportation Research Part A,* Vol. 41, No. 10, 2007, pp. 899–912.
18. Daly, A., and C. Rohr. Forecasting Demand for New Travel Alternatives. In *Theoretical Foundations for Travel Choice Modelling* (T. Gärling, T. Laitila, and K. Westin, eds.), Pergamon, Amsterdam, Netherlands, 1998.
19. Cherchi, E., and J. de D. Ortúzar. On Fitting Mode Specific Constants in the Presence of New Options in RP-SP Models. *Transportation Research Part A,* Vol. 40, No. 1, 2006, pp. 1–18.
20. Hensher, D. A., and M. Bradley. Using Stated Response Data to Enrich Revealed Preference Discrete Choice Models. *Marketing Letters,* Vol. 4, No. 2, 1993, pp. 139–152.
21. Hensher, D. A. Empirical Approaches to Combining Revealed and Stated Preference Data: Some Recent Developments with Reference to Urban Mode Choice. *Research in Transportation Economics,* Vol. 23, No. 1, 2008, pp. 23–29.
22. Ortúzar, J. de D., and C. Simonetti. Modelling the Demand for Medium Distance Air Travel with the Mixed Data Estimation Method. *Journal of Air Transport Management,* Vol. 14, No. 6, pp. 297–303.
23. Hensher, D. A., and J. M. Rose. Development of Commuter and Non-Commuter Mode Choice Models for the Assessment of New Public Transport Infrastructure Projects: A Case Study. *Transportation Research Part A,* Vol. 41, No. 5, 2007, pp. 428–443.
24. Hensher, D. A., J. J. Louviere, and J. Swait. Combining Sources of Preference Data. *Journal of Econometrics,* Vol. 89, No. 1/2, 1999, pp. 197–221.
25. Williams, H. C. W. L. On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit. *Environment and Planning A,* Vol. 9, No. 3, 1997, pp. 285–344.
26. Carrasco, J. A., and J. de D. Ortúzar. Review and Assessment of the Nested Logit Model. *Transport Reviews,* Vol. 22, No. 2, 2002, pp. 197–218.
27. McFadden, D. Econometric Models of Probabilistic Choice. In *Structural Analysis of Discrete Data with Econometric Applications* (Manski and D. McFadden, eds.), Massachusetts Institute of Technology Press, Cambridge, Mass., 1981.
28. Daly, A. J. Alternative Tree Logit Models: Comments on a Paper by Koppelman and Wen. *Transportation Research Part B,* Vol. 35, No. 8, 2001, pp. 717–724.
29. Bierlaire, M. BIOGEME: A Free Package for the Estimation of Discrete Choice Models. Presented at 3rd Swiss Transport Research Conference, Monte Verit, Ascona, Switzerland, 2003.
30. Ortúzar, J. de D., and L. G. Willumsen. *Modelling Transport,* 3rd ed. John Wiley and Sons, Chichester, United Kingdom, 2001.

The Transportation Demand Forecasting Committee peer-reviewed this paper.

Review of Evidence for Temporal Transferability of Mode–Destination Models

James Fox and Stephane Hess

One main motivation for developing travel behavior models is to use them to forecast future levels of transport demand. Given that the interest in transport planning is often in long-term forecasts, with forecast horizons of up to 30 years, it is important to consider the transferability of travel behavior models over time. The importance of model transferability has been recognized since disaggregate models were first applied in the late 1970s and early 1980s, but seems to have been largely forgotten recently, because the focus has been on the development of ever more advanced models that better explain current behavior, with a particular focus on the representation of taste heterogeneity. However, there are sufficient grounds to suspect that the model that best explains current behavior may not necessarily be the best tool for forecasting, not least because of the risk of overfitting to the base data. This paper aims to return the crucial issue of temporal transferability of travel demand models to the research agenda. First, the notion of transferability is discussed, highlighting the potential impacts of violations of the assumption of transferability, and the way transferability can be assessed is also described. Next, the most complete review of existing work investigating the temporal transferability of mode and mode–destination models to date is presented. Finally, a number of areas in which future research should be directed are identified.

The main interest in the field of travel behavior research lies in the development of models that are able to closely replicate choices made by travelers in real-life settings. The development of these models has two main aims; the models are used to understand current travel behavior, and they are applied to generate forecasts of future behavior. For the former, efforts focus mainly on producing accurate measures of willingness-to-pay indicators, for example, for use in appraisal. For the latter, various different contexts arise, namely, to forecast behavior under different scenarios (e.g., new transport infrastructure), to apply a model developed in one area to another area, and to use a model to produce long-term forecasts of future behavior. Notwithstanding the possibility of all three playing a role at the same time, it is the last of these, namely, the long-term forecast, that is at the heart of the issues discussed in this paper.

The importance to transport practice of producing accurate temporal forecasts should not be underestimated. Such forecasts are used by local and national government agencies to give an indication of likely future demand for the provision of transport services, and they help shape policy decisions, for example, in the context of new infrastructure developments. The complexity of this process is further increased by the need to take account of demographic changes, as well as the impact of changes in the transport infrastructure.

To make these forecasts, the approach that is typically followed is to develop models that represent a tractable simplification of current behavior and then to use those models to forecast future behavior. This means that the more advanced types of models, such as mixed multinomial logit, are rarely used in this context because the computational requirements they impose in application are too great. Indeed, although this cost may be justified in estimation, application relies on running the model in a potentially very large number of different contexts, often iteratively. The forecasting problem is further simplified by separating the key travel choice decisions on a given day, typically,

- Travel frequency—whether to travel and, if so, how many times;
- Mode of travel;
- Destination zone; and
- Time of day.

For each of these choices, separate models are usually developed by travel purpose because experience has demonstrated that the factors influencing these choices vary according to travel purpose. The focus of this paper is on the mode and destination choice decisions, which may be modeled as sequential choices or as a simultaneous choice.

In a forecasting context, mode–destination models are used to assess the effectiveness of different policies over forecasting horizons of up to 30 years. These models can include detailed socioeconomic segmentation, enabling a better fit to the estimation data set and an ability to predict the impact of trends in the input variables over time, such as increasing car ownership or aging of the population.

However, forecasting with such models relies on a significant assumption, namely, that the parameters that describe behavior in the base year can be used to predict future behavior, an issue that is referred to as transferability. In recent years, this issue has dropped off the radar, with the majority of effort going into the development of ever more advanced models that better explain current behavior, with a particular focus on the representation of taste heterogeneity. However, it is possible that the model that best explains current behavior may in fact not necessarily be the best tool for forecasting, not least because of potential issues of overfitting.

The problem is that if the assumption of transferability is violated, the future forecasts will be subject to uncertainty, irrespective of how well the models fit in the base year, how much segmentation they

Institute for Transport Studies, University of Leeds, Leeds, United Kingdom, LS2 9JT. Corresponding author: J. Fox, tsjbf@leeds.ac.uk.

Transportation Research Record: Journal of the Transportation Research Board, No. 2175, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 74–83.
DOI: 10.3141/2175-09

incorporate, and how accurately future model inputs can be forecast. As reflected in the discussions in this paper, the topic of transferability has received less and less attention in recent years. So although the use of models in forecasting remains one of the two main aims of travel behavior research, that is not reflected in current research activity.

The issue of what is meant by transferability is explored further in the next section. For the purposes of this introduction, it is useful to cite Koppelman and Wilmot, who define a transfer as "the application of a model, information, or theory about behavior developed in one context to describe the corresponding behavior in another context" (*1*).

This paper is concerned with the transferability of models, rather than underlying behavioral theories, in the context of model forecasting. In forecasting, models developed at one point in time are applied to predict behavior at a future point in time. It is thus assumed that the models are temporally transferable, that is, that the parameters that explain travel behavior when the model was estimated will also explain future travel behavior.

The aim of this paper is to return the crucial issue of temporal transferability of travel demand models to the research agenda. In particular, this paper examines the evidence for the transferability of mode destination choice models, which are applied over forecasting horizons of up to 30 years, but with, as is demonstrated, little evidence for their transferability over such periods.

This paper has three main components. First, the notion of transferability is discussed, highlighting the potential impacts of violations of the assumption of transferability, and the way in which transferability can be assessed is described. This is followed by what the authors believe to be the most complete review of existing work in this area to date. Finally, a number of areas in which future research should be directed are identified.

TRANSFERABILITY

Defining Transferability

Koppelman and Wilmot offer the following definition of transferability, which is, in the authors' view, the best definition provided in the literature:

> First, we define transfer as the application of a model, information, or theory about behavior developed in one context to describe the corresponding behavior in another context. We further define transferability as the usefulness of the transferred model, information or theory in the new context. (*1*)

The first part of this definition can be interpreted quite broadly. For example, applying a model based on principles of utility maximization assumes that those principles apply in the context in which the model is applied, as well as in the context in which the model is developed. However, the focus of the transferability literature, and this paper, is on model transferability. That is to say, assessing the ability of models developed in one context to explain behavior in another context, under the assumption that the underlying behavioral theory on which the model is based is equally applicable in the two contexts.

Somewhat surprisingly, none of the other papers reviewed attempted to set out their own definition of transferability, and indeed in many cases the term is used under the assumption that its meaning is known to the reader.

Temporal and Spatial Transferability

A key distinction made in the literature is between temporal transferability and spatial transferability. Temporal transferability is concerned with the application of models developed by using data collected at one point in time at another point in time, whereas spatial transferability is concerned with the application of models developed using data from one spatial area in another spatial area. Usually temporal transfers take place in the same spatial area, and spatial transfers take place at or around the same point in time. However, in some cases a model is transferred over both time and space, and so the two categories are not mutually exclusive.

To consider temporal and spatial transferability in the context of disaggregate mode destination choice models, it is useful to define in summary form the utility functions used in the models:

$$U_{md} = \beta \cdot X + \epsilon_{md} \quad (1)$$

where

U_{md} = utility of mode–destination alternative md,
β = vector of model parameters,
X = vector of observed data, and
ϵ_{md} = random error term.

In model development, the objective is to identify model parameters that best explain the observed data. Thus, as a model is developed, and its ability to explain the observed choices increases, the term X increases in importance, and the term ϵ_{md} decreases in importance. Nonetheless, mode destination models do not perfectly explain the observed choices, and so some random error remains. The mean effect of this term is captured in the mode-specific constants, which in a mode choice context will capture effects such as the relative reliability of modes, levels of comfort, climate, and hilliness for walking and cycling.

In a spatial transfer at the same point in time, the transferability of the model will depend on the relevance of the parameters in the transfer context, for example, the degree of similarity in sensitivities to travel time and cost, and on the appropriateness of the alternative specific constants. Models would be expected to be transferable for areas that have similar characteristics, such as similarities in mean travel times and costs, levels of highway and public transport reliability, climate, and hilliness.

For a temporal transfer in a given area, the considerations are different. The effect of area-to-area differences is not present, instead the key issue is whether the parameters remain constant over time. Stated more explicitly, the issue is whether within a given population segment, sensitivities to the different variables that form the utility functions, and the mean contribution of unmeasured effects as measured by the alternative specific constants, remain constant over time. In some instances, the ratio between model parameters is also important; for example, the value of time implied by the ratio between the cost and time parameters in a model, which will change over time if there are changes in the cost and time parameters.

Thus temporal transferability and spatial transferability are not the same thing. A model might be temporally transferable within a given area, but contain a specification that does not transfer well to other areas. Another model might contain a detailed specification that transfers well to other spatial areas, but does not transfer well over time.

Spatial transfers typically involve a transfer sample, a sample of choices observed in the transfer context, which may allow a locally

estimated model to be developed for comparison with the model transfer. When a model is applied to forecast future behavior, that is a transfer of the model to a new temporal context. However, unlike many spatial transfers, no transfer sample is available. There is, therefore, an important practical difference between temporal and spatial transfers.

Temporal transferability can be assessed, however, by using two data sets collected at different points in time from the same spatial area. Typically one data set is historical, one is contemporary. Models estimated from the two samples can be compared to make assessments of model transferability, and from these, conclusions can be drawn about the temporal transferability of similar models used for forecasting. Such an assessment, however, also needs to take into account that although models may be temporally transferable in one area, that may not be the case in another area or in a different context. From that perspective, such assessments need to be based on the use of contexts that are as similar as possible to the study context.

This paper is concerned with the temporal transferability of mode destination models during long-term forecasting horizons. It is worth emphasizing that during such forecasting horizons, key model inputs, such as population, employment, and travel times and costs on the networks, will be subject to considerable uncertainty, and different assumptions can have substantial effects on the predictions of future travel behavior. For example, a perfectly transferable model might be fed with poor predictions of input variables and consequently produce poor quality forecasts. Thus, as illustrated later in this paper, temporal transferability is a factor in producing the best possible forecasts of future behavior, but is certainly not the only consideration.

Conditions for Transferability

A theme in a number of the early papers on the transferability of disaggregate models is a belief that disaggregate models, which represent choice at the individual level, should be more transferable than aggregate models, which typically represent choices at the zonal level. In some cases, claims were made for the models without much supporting evidence. For example, Ben-Akiva and Atherton claimed, in the context of spatial transferability, that

> a second major advantage of the disaggregate demand modelling approach is that it is transferable from one urban area to any another. It has been hypothesised that, because disaggregate models are based on household or individual information and do not depend on any specific zone system, their coefficients should be transferable between different urban areas. (*2*)

Although the second sentence of this quote concedes that transferability is a hypothesis, the first seems to treat it as a given for a transfer to any area. The argument about the zone system seems to have been made in reference to aggregate modeling approaches, which typically operate at the zonal level, but the arguments were not identified. More generally, although a number of these early papers in the transferability literature claim that disaggregate models are more transferable than aggregate techniques, only Watson and Westin empirically demonstrated that claim (*3*).

Later works, building on empirical findings that the disaggregate models were not always transferable, were more measured in their claims. Daly identified three conditions for spatial transferability (*4*):

- Relevance. Does the local model give any information on travel behavior in the transfer area?
- Validity. Is the transfer model acceptably specified for the transfer area?
- Appropriateness. Is it appropriate to use the transferred model in the target area?

Thus models are expected to be transferable only under certain circumstances. An important finding from the literature review was that although some authors had attempted to identify conditions for spatial transferability, to the present authors' knowledge no corresponding research exists for temporal transferability.

Impacts of Violation of Transferability Assumption

If temporal transferability does not hold, what are the implications for forecasting? When the model is used to forecast future behavior, the forecasts will be subject to error as a result of differences between the model parameters and the true model parameters in the future year. This will add error into the forecasts, and the magnitude of this error would be expected to be larger the longer the forecast period, that is, the longer the interval during which the model is transferred. Temporal transferability is not stated here as the only condition that must be satisfied to produce accurate forecasts, rather it is a factor that is often neglected. By contrast, significant effort may go into predicting the composition of the future population and other model inputs.

Figures 1 and 2 seek to illustrate that point. In this simple example, the base car share in 2000 is 40%, which is forecast to grow steadily to 70% by 2030. However, as a result of uncertainty in the input variables, the uncertainty in this prediction is ±10%. In the first figure, the model is taken to be perfectly transferable, and therefore the overall uncertainty in the 70% mode share for car is ±10%. In the second example, uncertainty due to input variables is again taken to be ±10%, but uncertainty due to the transferability of the model is also ±10%. Thus, the overall uncertainty in the forecasts is ±20%. What these figures aim to illustrate is that model transferability may add further uncertainty to model forecasts. One approach modelers use to deal with uncertainty in the future input variables is to run models for different scenarios, for example, by running low-, medium-, and high-growth scenarios. However, understanding the uncertainty introduced into forecasting by the input variables and model transferability would give a more complete picture of the true levels of uncertainty associated with future forecasts and the relative importance of these two effects.

Assessing Transferability

In a temporal forecast context, testing for transferability is not possible in advance. Indeed, forecasts for a future period are being produced, and the accuracy of these forecasts can be assessed only in the future. Evidence on the temporal transferability of particular types of models can, however, be produced by looking at historical studies, that is, studies in which one is in position to compare the forecasts to what actually occurred in reality. Specifically, temporal transferability can be assessed by using data sets that have been collected at two points in time in the same geographical area. Provided identical, or similar, variables are collected in the two cases, it is possible to use the sets of data to develop identically specified models at both points in time and make assessments of model transferability. This generally makes the assumption that the actual model

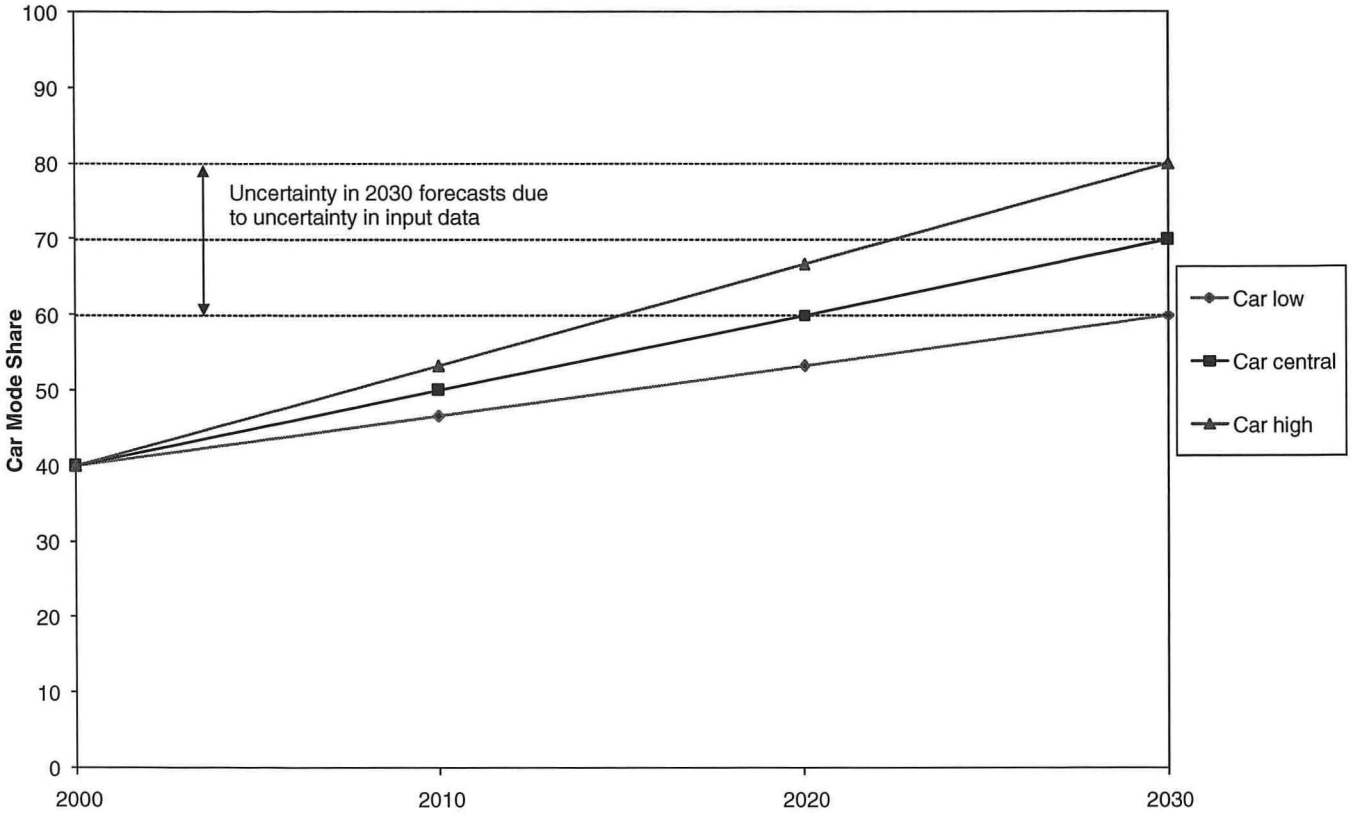


FIGURE 1 Uncertainty in model inputs only.

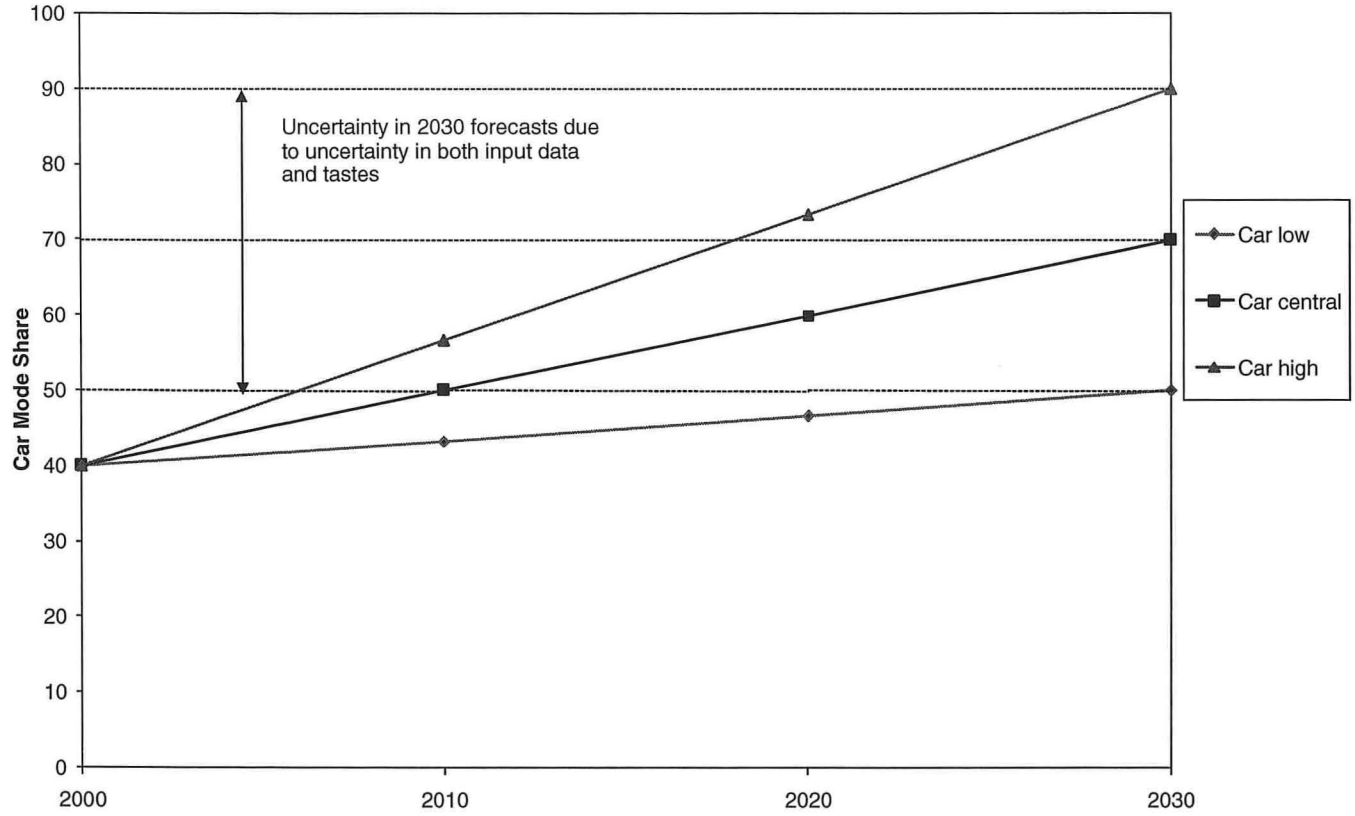


FIGURE 2 Uncertainty in model inputs and model parameters.

type is transferable and that transferability is influenced only by the specification of the utility function. This assumption is based largely on the fact that in applied work, generally the same basic model structures are used throughout, but the possibility that different model structures may be more appropriate at different times, and the impact of this on forecasting abilities, is an interesting area for future research.

The measures of transferability used in the literature can be placed into two broad categories. First are tests of parameter equality. These represent strict statistical tests of the hypothesis of parameter transferability and were the key measures of transferability used in the early literature. Many of these tests rely on the availability of a transfer sample, which is used to develop a locally estimated model, and then the transferred model is assessed relative to that locally estimated model.

The second category is predictive measures, which are assessments of the predictive ability of a model in the transfer context. Predictive measures can be used to make assessments of model transferability, but they do not necessarily directly measure transferability and so need to be interpreted with caution. They are, however, arguably less reliant on the assumption that the same model structure applies in both contexts.

Tests of Parameter Equality

A frequently used statistical test in the literature is the transferability test statistic (TTS), which assesses the transferability of the base model parameters b in the transfer context t, under the hypothesis that the two sets of parameters are equal:

$$\mathrm{TTS}_t(\beta_b) = -2\left[\mathrm{LL}_t(\beta_b) - \mathrm{LL}_t(\beta_t)\right] \tag{2}$$

where $\mathrm{LL}_t(\beta_b)$ is for the base model applied to the transfer data and $\mathrm{LL}_t(\beta_t)$ is for the locally estimated model.

TTS is χ^2 distributed with degrees of freedom equal to the number of model parameters. It can be seen that this test is the same as the standard likelihood ratio test but applied to pairs of log likelihood values in a different context.

The transfer index (TI) was devised by Koppelman and Wilmot and measures the predictive accuracy of the transferred model relative to a locally estimated model, with an upper bound of one (*1*). A reference model is used in the calculation of TI, typically a market shares model in the case of mode choice.

$$\mathrm{TI}_t(\beta_b) = \frac{\left(\mathrm{LL}_t(\beta_b) - \mathrm{LL}_t\left(\beta_t^{\mathrm{ref}}\right)\right)}{\left(\mathrm{LL}_t(\beta_t) - \mathrm{LL}_t\left(\beta_t^{\mathrm{ref}}\right)\right)} \tag{3}$$

where β_t^{ref} is the reference model for the transfer data and $\mathrm{LL}_t(\beta_t) \geq \mathrm{LL}_t(\beta_b) \geq \mathrm{LL}_t(\beta'_{\mathrm{ref}})$.

Unlike the TTS, the TI does not either accept or reject the hypothesis of model transferability. Rather it provides a relative measure of model transferability. Within a given study area, the TI can be used to directly assess different sets of models. When comparisons between different studies are being made, the TI still provides insight if the same reference model specification is used but does not have a general scale in a formal sense.

The statistical measures discussed above are concerned with the overall fit to the data and are the measures that have been used in the literature to assess transferability. It is also possible to analyze differences in individual parameter values by using information on the significance of the parameter in the base and transfer models. For example, the cost and time parameters in a model are key to the forecast responses to policy, and so changes in these parameters over time are of particular relevance.

Predictive Measures

As was discussed in the introduction to this section, predictive measures were increasingly used to assess transferability as the transferability literature developed. For example, Lerman argued that the early transferability literature had used an overrestrictive definition of transferability, with an overemphasis on statistical tests, and argued that transferability should not be seen as a binary issue, but rather that the extent of transferability should be explored (*5*). In the same book, Ben-Akiva argued that achieving perfect transferability is impossible because a model is never perfectly specified, and therefore pragmatic transferability criteria are required in addition to standard statistical tests (*6*).

Predictive measures need to be interpreted carefully when one is making assessments of model transferability. In cases in which both base and transfer samples are available, then provided both data sets provide accurate samples of individual choices, the ability of the base model to predict choices in the transfer context is a direct test of the transferability of the model.

However, in many studies that validate model predictions against observed outcomes, a detailed transfer sample is not available, and the model forecasts are validated against aggregate shares. In these studies the predictions of the model depend on the accuracy of the assumed inputs as well as the transferability of the model itself. So, a model may be highly transferable, but if fuel prices dramatically increase during the forecast period, and that was not anticipated when the future inputs were assembled, the model predictions may be some way off the observed outcomes. Care needs to be taken to distinguish input errors from transferability errors, and in some cases it is not possible to disentangle the two effects.

The relative error measure (REM) has been used in the literature to assess model transferability. It assesses for the prediction the ability of a model to predict the choice frequency in some aggregate group as follows:

$$\mathrm{REM}_{mg} = \frac{\left(P_{mg} - O_{mg}\right)}{O_{mg}} \tag{4}$$

where P_{mg} is the prediction for alternative m in group g and O_{mg} is the observed choices for alternative m in group g.

Note that g is often dropped, that is, predicted and observed alternative (e.g., mode) shares are compared. Because the REM measure is self-scaling, it can be applied to probabilities and to aggregate choice predictions such as numbers of individuals choosing m and g.

LITERATURE REVIEW

The literature on temporal transferability has been broken down into three subsections. The first two discuss studies using disaggregate mode choice models and thus are more directly relevant than is the other literature to the focus of this paper on models of mode and destination choice. The final subsection then presents evidence from other model types, in most cases aggregate models of trip generation.

The mode choice studies are further broken down into direct tests of model transferability, in which both base and transfer models have

been developed allowing formal statistical tests of transferability to be made, and validation studies, in which model predictions are compared with aggregate statistics on mode share, often after substantial changes to travel times, costs, or both. These validation studies use data collected in the transfer context to define the inputs to the models, which removes the complication of combinations of errors in the input data discussed in the section on predictive measures. A number of the papers present comparisons of base and transfer models and use the transfer data to validate the performance of the base model in forecasting, and so are discussed in both subsections.

Mode Choice Transferability Studies

Four studies of the transferability of mode choice models have been reviewed. Train compared models developed before (in 1972) and after (in 1975) the opening of the Bay Area Rapid Transit (BART) system in San Francisco, California (*7*). Silman compared a model for Tel Aviv between 1972 and 1976 (*8*). McCarthy also analyzed pre-BART data, in his case from 1973–1974, with the post-BART data from 1975 (*9*). Badoe and Miller developed models from two large household interviews collected 22 years apart (*10*). All four studies analyzed home–work trips only.

In addition to these four mode choice studies, two studies have investigated the transferability of models of simultaneous mode and destination choice, the exact focus of this paper. Karasmaa and Pursula (*11*) used Helsinki, Finland, data from 1981 and 1988, and Gunn (*12*) investigated models for the Netherlands by using 1982 and 1995 data. As did the four mode choice studies, Karasmaa looked at home–work trips only, but Gunn ran analyses for home–work, home–shopping, and home–social and home–recreational.

Overall, the mode choice studies supported the hypothesis that model parameters are reasonably stable over time, although this finding was not universal; two of the six studies reported substantial changes over time. Silman and McCarthy both used the TTS and were able to accept the hypothesis of temporal parameter stability at a 10% confidence interval, although McCarthy rejected the hypothesis at a 5% confidence level. The Badoe and Miller study is noteworthy; it is the only study that considers a long-term forecasting interval. Badoe and Miller rejected the hypothesis that the parameters were equal during a 22-year period, but for some model specifications TI values of almost 0.9 were obtained. Thus a transferred model from 1964 used to predict 1986 behavior had 90% of the predictive ability of a local model estimated on 1986 data.

Neither of the mode–destination studies (Karasmaa, Gunn) calculated TTS or TI values. Gunn's findings of general parameter stability were consistent with the mode choice studies; however, in Karasmaa's analysis there were significant differences between the base and transfer parameters.

Badoe and Miller made an interesting assessment of the impact of model specification on model transferability by testing seven different model specifications, ranging from simple market shares models and models with mode constants and level-of-service variables only, through to models with detailed market segmentation. For all model specifications, the TTS rejected the hypothesis of parameter stability at a 5% confidence interval. The TI increased from 0.132 for the simple market shares model, to 0.894 in the level-of-service variables only model, although interestingly more detailed specifications with market segmentation had lower TI values, despite higher log likelihood values.

This finding raises an interesting question as to whether, for long-term forecasting, there is an optimum level of complexity to ensure that the predictive ability of the model is retained over time. It may be that adding detailed market segmentations improves the fit to the base data, but that this is a case of overfitting and gives less robust forecasts over the longer term.

Mode Choice Validation Studies

Parody assessed the impact, during a 1-year period, of a free bus service accompanied by substantial increases in parking changes at the University of Massachusetts at Amherst (*13*). Ben-Akiva and Atherton predicted the impact of preferential lanes on bus usage and carpooling along the Shirley Highway in Washington, D.C. (*2*). Train (*7*) validated the ability of pre-BART models to predict demand when BART was introduced and then investigated how the forecasting performance of the models varied with model specification in Train (*14*). Silman used a model for Tel Aviv, Israel, developed with 1972 data to predict behavior in 1976 (*8*). Milthorpe's study had a different focus, providing a comparison of the forecasts of a four-stage model developed in the early 1970s with observed data from about 2001 (*15*).

The general pattern from these studies is that the mode choice models were able to predict the impact of often substantial changes in level of service on mode share with reasonable accuracy. This finding is reassuring for the application of mode choice models for periods of up to 5 years, but it does not provide any direct evidence about the transferability of the models during the longer term.

Parody's analysis used panel data and in one test assessed the impact of substantial increases in parking charges. In this test, a full model specification with socioeconomic parameters performed substantially better than a model with level-of-service parameters alone. This suggests that an improved model specification yielded more transferable level-of-service parameters. Train's 1979 analysis also concluded that improving the model specification resulted in improvements in the model predictions.

It seems that the improvement in the predictive performance of the models that results from adding socioeconomic parameters is a result of improved estimates of the key level-of-service parameters, rather than the impact of changes in socioeconomics, given that most of these model tests have been undertaken over short-term forecasting horizons of up to 5 years. These improved estimates then enable the models to better predict the impact of changes in level of service. Silman explicitly noted that pattern by observing that when socioeconomic parameters were added, the significance of the key cost and time variables in his models was improved.

Following from the discussion above of the danger of overfitting to the base data, there is clearly a need to find the appropriate balance in regard to the level of detail in the model. Adding socioeconomic parameters has been found to improve the estimates of the core level-of-service parameters; however, there is a danger that adding too much detail leads to overfitting and less robust forecasts during the longer term. Further empirical analysis of this issue would be valuable.

Other Transferability Studies

A number of other studies provide insight into the temporal transferability of models. The following paragraph summarizes the various papers reviewed, and following that there is a discussion of the findings.

Hill and Dodd used 1956 and 1964 household interviews from Toronto, Ontario, Canada, to assess zonal regression trip generation models (*16*). Kannel and Heathington investigated the stability of household regression trip generation models estimated from 1964 and 1971 household surveys in Indianapolis, Indiana (*17*). Downes and Gyenes compared the predictive performance of three trip generation techniques, zonal regression, category analysis, and household regression, by using data from Reading, United Kingdom, collected in 1962 and 1971 (*18*). Yunker analyzed the predictive performance of trip generation and distribution models by using 1963 and 1972 data from Wisconsin (*19*). Smith and Cleveland investigated the time stability of household regression trip generation models for Detroit, Michigan, by using household interviews from 1953 and 1965 (*20*). Doubleday analyzed the same set of Reading data as did Downes to investigate the temporal stability of category analysis trip generation models (*21*). Elmi et al. used Toronto data to investigate the temporal stability of aggregate trip distribution models by using data from 1964, 1986, and 1996 (*22*). Cotrus et al. investigated the transferability of trip generation demand models by using data from the 1984 and the1996 to 1997 Israeli National Travel Habits survey (*23*).

Most of these studies are concerned with generation modeling and typically used aggregate modeling approaches based on regression, household classification, and gravity model techniques. As such, any findings with respect to model transferability have to be interpreted with caution for the mode–destination modeling context. Nonetheless, general findings are of interest to the broader question of whether models developed at one point in time can be used to predict behavior at a future point in time. These studies also have the advantage that they have tended to consider longer forecasting intervals, typically about 10 years, compared with the mode choice studies.

Few of these studies made formal statistical tests of model transferability. Elmi concluded that the parameters in his trip distribution models were statistically different between 1964 and 1986, although the 1964 models were able to predict 1986 behavior well. Cotrus also rejected the hypothesis of temporal stability, in Haifa and in Tel Aviv, Israel, during a 12- to 13-year period.

The assessments of the predictive performance of the generation models are supportive of the hypothesis of model transferability, with five of the seven studies reporting that the models predicted future trip generations well. However, as discussed earlier, accurate aggregate predictions do not necessarily indicate transferability at the individual parameter level.

A noteworthy feature of many tests of the generation models is that the intervals of analysis often covered substantial changes in population, whereas the mode choice validation studies typically were concerned with the effect of substantial changes in travel cost and times. For example, Hill and Dodd's analysis covered a period when the population of the Greater Toronto area increased by 33% and total car ownership rose by 45%. The good predictive performance of the models under these conditions provides some evidence for the temporal stability of socioeconomic parameters that capture variation in behavior across the population.

Elmi's analysis of work trip distribution models investigated the impact of improving the model specification and, consistent with the mode choice studies, he concluded that improved model specification resulted in improved model transferability. Elmi obtained TI values as high as 0.84 for predicting 1996 behavior with 1964 models, and 0.97 for predicting 1996 behavior with 1986 models. An interesting result noted by Elmi was that the disutility of travel time reduced over time, from a value of –0.13 in 1964 to –0.08 in 1996. Elmi suggested that this reflected changes in spatial structure and consequent increases in the mean distance to work.

Elmi's hypothesis that changes in model parameters might be related to changes in spatial structure may give an approach for forecasting how model parameters change over time. If evidence were assembled across studies of the way model parameters had changed over time, it would be possible to investigate whether the changes in model parameters could be explained in relation to aggregate variables describing changes in spatial structure, such as the size of the urbanized area.

Summary and Critique

To draw the findings from the review of temporal transferability together, it is useful to summarize the key findings from the groups of studies. These summaries are presented in Tables 1 to 3.

Overall, the direct tests of transferability summarized in Table 1 are supportive of the hypothesis that mode choice models can be transferred over time, with four of the six studies concluding that the models tested were transferable. Furthermore, some of the validation studies demonstrate that the models are able to predict the impact on mode share of substantial changes in level of service during short periods.

That said, these findings are specific to the evidence base that has been analyzed. Considering the direct tests of temporal transferability summarized in Table 1, it can be seen that the evidence is nearly all from commuting studies. Furthermore, all the validation studies summarized in the second table, and many of the generation studies summarized in Table 3, are also based on commuter travel. Commuting travel might be expected to be more transferable than other purposes because the journey to work is a regular trip and as such would be expected to be recorded with a higher degree of accuracy than less regular trips.

Another feature of the evidence base is that much of it is based on short-term forecasts of up to 10 years. However, many transport models are applied over forecast periods of up to 30 years, and it seems reasonable to hypothesize that during longer time intervals, transferability would be less likely to be accepted. That said, the single body of evidence on longer term transferability, the studies from Toronto that developed mode choice models and distribution models, is supportive of model transferability.

An empirical finding from both mode choice and distribution studies is that improving model specification improves model transferability. Although the improvements in model specification described are often the addition of socioeconomic parameters, this improvement in model performance seems to come about because the improved models provide better estimates of the key cost and time parameters that respond to short-term policy changes. During a longer term forecasting horizon, substantial changes in the distribution of the population across segments would be expected, and so the findings in regard to model specification may be different, depending on the relative stability of level-of-service and socioeconomic parameters during the longer term.

Only two studies of temporal transferability have considered simultaneous models of mode and destination choice, the focus of this particular paper. Gunn's study found a good level of temporal transferability, but in Karasmaa's work three out of four level-of-service parameters were not transferable.

The dates of the studies are noteworthy, with half (nine out of 18) published in the 1970s, and with only two papers published during the past decade. Clearly research efforts into the issue of model transferability have been limited since the cluster of work in the 1970s

TABLE 1 Temporal Mode Choice Transferability Studies

Paper and Reference	Area	Purpose	Time Frame	Degree of Transferability	Comments
Train (1978) (*7*)	San Francisco	Commute		LOS parameters more stable than other terms	
Silman (1981) (*8*)	Tel Aviv	Commute	4 years (1972–1976)	Good—time parameters particularly stable	
McCarthy (1982) (*9*)	San Francisco	Commute	1.5 years (1973, 1974–1975)	Parameters stable over short term	Box–Cox transforms used
Badoe and Miller (1995) (*10*)	Toronto	Commuter mode choice	22 years (1964–1986)	Statistical differences between parameters but models are broadly transferable in terms of predictive performance; ASCs and scale change over time	
Karasmaa and Pursula (1997) (*11*)	Helsinki	Commute	7 years (1981–1988)	Poor—three of four LOS parameters not stable	Mode–destination models
Gunn (2001) (*12*)	Netherlands	Commute, personal business, shopping, social, and recreational	13 years (1982–1995)	Good, particularly for LOS parameters	Mode–destination models, some evidence that transferability may vary with purpose

NOTE: LOS = level of service; ASC = alternative specific constant

and early 1980s. In addition, the evidence that models of mode–destination choice are temporally transferable during forecasting intervals of up to 30 years is extremely limited. Given the importance of such long-term forecasts in transport planning, this is a serious shortcoming in the field and an important area for future research.

DIRECTIONS FOR FURTHER RESEARCH

It is clear from this review that further empirical evidence on the temporal transferability of mode–destination models would be valuable and, in particular, give insight into their suitability for forecasting during longer term forecasting horizons.

Comparisons between model predictions and observed aggregate outcomes can provide valuable insight. However, it is difficult to disentangle the impact of model transferability from other factors, as a result of, in particular, the magnitude of errors in input assumptions during long-term horizons. For example, Milthorpe analyzed forecasts for Sydney, Australia, from 1971 to 2001 (*15*). During this period, the population was predicted to grow by 55%, but actually grew by 35%. Clearly such errors have a substantial impact on the model predictions, irrespective of the transferability of the models.

The best approach for future research is to focus on cases in which detailed disaggregate data, such as household interview data, are available during periods of 20 to 30 years. Provided that reasonable levels of consistency exist between the data collected at each point in time and that consistent level-of-service data can be assembled for each point in time, such data sets can be used to directly test the assumption of temporal transferability.

TABLE 2 Temporal Mode Choice Validation Studies

Paper and Reference	Area	Purpose	Time Frame	Predictive Performance	Comments
Parody (1977) (*13*)	University of Massachusetts at Amherst	Commute	Four waves: 1. Autumn 1972 2. Spring 1973 3. Autumn 1973 4. Spring 1974	Good—substantial improvement when model specification improved with socioeconomic terms	Large changes in modal costs over time period
Ben-Akiva and Atherton (1977) (*2*)	Washington, D.C., and Santa Monica, Calif., United States (application only)	Commute	Washington, D.C. 1970–74 Santa Monica, 1974	Good in response to significant changes in LOS	Focus on carpooling policies
Train (1978, 1979) (*7, 14*)	San Francisco	Commute		Poor for transit because of problems with input data; predictions improve with improved model specification	Lack of info. for new BART mode, erroneous walk time data
Silman (1981) (*8*)	Tel Aviv	Commute	4 years (1972–1976)	Mixed—main car driver and bus modes predicted well; minor car passenger mode significantly overpredicted	

TABLE 3 Temporal Generation Model Studies

Paper and Reference	Area	Model Class	Purpose(s)	Time Frame	Evidence for Transferability?	Comments
Hill and Dodd (1966) (*16*)	Toronto	Zonal regression	All purposes, all purposes peak hour	8 years (1956–1964)	Yes, after correcting for differences in data processing	Actual results after correction applied unclear
Kannel and Heathington (1973) (*17*)	Indianapolis	Household regression	All purposes	7 years (1964–1971)	Yes, predicted trips within 2% of observed	Panel of households used, this may have influenced findings
Downes and Gyenes (1976) (*18*)	Reading	Zonal regression, category analysis, household regression	All purposes, plus split into shop, work, other	9 years (1962–1971)	Yes, forecasting errors close to base year errors	
Yunker (1976) (*19*)	Southeast Wisconsin, United States	Zonal regression analysis	Commute, shopping, other, non-home-based	9 years (1963–1972)	Good, predicted growth close to observed, larger differences by purpose	Observed trips grew by 25% in period
Smith and Cleveland (1976) (*20*)	Detroit	Category analysis, household regression	All purposes	12 years (1953–1965)	No, trip rates not stable, uniform growth over categories	Uniform growth likely to be income and/or accessibility
Doubleday (1977) (*21*)	Reading	Aggregate, category analysis	Regular (work) and nonregular	9 years (1962–1971)	Trip rates not stable, exception was employed males	Accessibility had an impact, and possibly income growth
Cotrus, Prashker, and Shiftan (2005) (*23*)	Haifa and Tel Aviv	Person-level regression	All purposes	12–13 years (1984–1996, 1997)	Mixed, statistically rejected, but predictions good with 7% and 3% errors	

Three additional areas for future research are identified. The first of these goes back to the earlier point that even in studies testing for transferability, the assumption is generally made that the actual model type is transferable and that transferability is influenced only by the specification of the utility function. Here, future work should look at the transferability of the actual model form in addition to the specification. Second, the existing literature focuses almost exclusively on the transferability of the most simplistic types of models, generally multinomial logit. Currently, nearly all forecasting models make use of these simpler model forms, given the earlier point about the complexity and computational cost of more advanced models. However, it seems that an interesting avenue for research in this context would be to test whether the use of advanced models, such as mixed multinomial logit, would improve transferability, which would make the higher computational cost more acceptable. Finally, although implicit in the background of most work, developing and applying tests for model transferability is clearly only a means to an end, and the overarching aim of future research should be to provide guidance on how the transferability of models can be improved.

ACKNOWLEDGMENTS

The authors acknowledge the input of Andrew Daly into the wider study of which this paper is a part. The first author acknowledges the financial support of a doctoral training award funded by the Engineering and Physical Sciences Research Council. The second author acknowledges the financial support of the Leverhulme Trust in the form of a Leverhulme early career fellowship.

REFERENCES

1. Koppelman, F., and C. Wilmot. Transferability Analysis of Disaggregate Choice Models. In *Transportation Research Record 895,* TRB, National Research Council, Washington, D.C., 1982, pp. 18–24.
2. Ben-Akiva, M., and T. Atherton. Methodology for Short-Range Travel Demand Predictions. *Journal of Transport Economics and Policy,* Vol. 11, 1977, pp. 224–261.
3. Watson, P., and R. Westin. Transferability of Disaggregate Mode Choice Models. *Regional Science and Urban Economics,* Vol. 5, 1975, pp. 227–249.
4. Daly, A. *A Study of Transferability of Disaggregate Mode Choice Models from Grenoble to Nantes.* Cambridge Systematics Europe, the Hague, Netherlands, 1985.
5. Lerman, R. A Comment on Interspatial, Intraspatial, and Temporal Transferability. In *New Horizons in Travel-Behavior Research* (A. M. Stopher, A. H. Meyburg, and W. Brög, eds.), Lexington Books, Lanham, Md., 1981, pp. 628–632.
6. Ben-Akiva, M. Issues in Transferring and Updating Travel-Behavior Models. In *New Horizons in Travel-Behavior Research* (A. M. Stopher, A. H. Meyburg, and W. Brög, eds.), Lexington Books, Lanham, Md., 1981, pp. 665–686.
7. Train, K. A Validation Test of a Disaggregate Mode Choice Model. *Transportation Research,* Vol. 12, 1978, pp. 167–174.
8. Silman, L. The Time Stability of a Modal-Split Model for Tel Aviv. *Environment and Planning A,* Vol. 13, No. 6, 1981, pp. 751–762.

9. McCarthy, P. Further Evidence on the Temporal Stability of Disaggregate Travel Demand Models. *Transportation Research B,* Vol. 16, No. 4, 1982, pp. 263–278.
10. Badoe, D. A., and E. J. Miller. Analysis of the Temporal Transferability of Disaggregate Work Trip Mode Choice Models. In *Transportation Research Record 1493,* TRB, National Research Council, Washington, D.C., 1995, pp. 1–11.
11. Karasmaa, N., and M. Pursula. Empirical Studies of Transferability of Helsinki Metropolitan Area Travel Forecasting Models. In *Transportation Research Record 1607,* TRB, National Research Council, Washington, D.C., 1997, pp. 38–44.
12. Gunn, H. Spatial and Temporal Transferability of Relationships Between Travel Demand, Trip Cost and Travel Time. *Transportation Research Part E,* Vol. 37, No. 2–3, 2001, pp. 163–189.
13. Parody, T. E. Analysis of Predictive Qualities of Disaggregate Modal-Choice Models. In *Transportation Research Record 637,* TRB, National Research Council, Washington, D.C., 1977, pp. 51–57.
14. Train, K. A Comparison of the Predictive Ability of Mode Choice Models with Various Levels of Complexity. *Transportation Research Part A,* Vol. 13, No. 1, 1979, pp. 11–16.
15. Milthorpe, F. A Comparison of Long Term Sydney Forecasts with Actual Outcomes. *Australasian Transport Research Forum,* Vol. 28, 2008.
16. Dodd, N. Studies of Trends of Travel Between 1954 and 1964 in a Large Metropolitan Area. In *Highway Research Record 141,* HRB, National Research Council, Washington, D.C., 1966, pp. 1–23.
17. Kannel, E. J., and K. W. Heathington. Temporal Stability of Trip Generation Relations. In *Highway Research Record 472,* HRB, National Research Council, Washington, D.C., 1973, pp. 17–27.
18. Downes, J., and L. Gyenes. *Temporal Stability and Forecasting Ability of Trip Generation Models in Reading.* Transport and Road Research Laboratory, Crowthorne, United Kingdom, 1976.
19. Yunker, K. R. Tests of the Temporal Stability of Travel Simulation Models in Southeastern Wisconsin. In *Transportation Research Record 610,* TRB, National Research Council, Washington, D.C., 1976, pp. 1–5.
20. Smith, R. L., Jr. and D. E. Cleveland. Time-Stability Analysis of Trip-Generation and Predistribution Modal-Choice Models. In *Transportation Research Record 569,* TRB, National Research Council, Washington, D.C., 1976, pp. 76–86.
21. Doubleday, C. Some Studies of the Temporal Stability of Person Trip Generation Models. *Transportation Research,* Vol. 11, 1977, pp. 255–263.
22. Elmi, A. M., D. A. Badoe, and E. J. Miller. Transferability Analysis of Work-Trip-Distribution Models. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1676,* TRB, National Research Council, Washington, D.C., 1997, pp. 169–176.
23. Cotrus, A., J. Prashker, and Y. Shiftan. Spatial and Temporal Transferability of Trip Generation Demand Models in Israel. *Journal of Transportation and Statistics,* Vol. 8, No. 1, 2005.

The Transportation Demand Forecasting Committee peer-reviewed this paper.

Patronage Ramp-Up Analysis Model Using a Heuristic *F*-Test

Justin S. Chang, Sung-bong Chung, Kyu-hwa Jung, and Ki-min Kim

This paper deals with a novel patronage ramp-up analysis model. The ramp-up effect represents the delay in ridership take-up during the first periods of new transport services or facilities. This influence is known as one of the external errors of transport demand forecasts, which degrades the reliability of demand studies and economic appraisal for transport schemes. Traditionally, professional judgment and regression analyses have been used to investigate this effect, but the existing methodologies suffer from the inherent limitations of being biased and arbitrary. In this paper, a heuristic *F*-test is suggested as an alternative framework. The model proposed is tested with road and rail schemes of South Korea. The case studies illustrate the ability of the model to find the ramp-up parameters involving the duration and the ratio of this effect. Subsequently, a numerical example of cost–benefit analysis including ridership ramp-up is provided. The experiment shows that incorporating the ramp-up effect could provide help for more cautious decision making during the appraisal. A summary and future directions of the study are also provided.

There are growing concerns with errors in transport demand forecasts (*1–4*). Three types of errors are usually identified. First, measurement or data errors are related to an insufficient data set to produce accurate descriptions of the existing transport system; the data set normally consists of demand-side characteristics (traffic patterns and volumes), supply profiles (travel times, journey distances, and frequencies and routes of public transport services), and evaluation parameters (annualization indices, factoring single-day observations to an average daily level, and working–nonworking time split). Because a complete data set is costly to obtain, studies rely on estimation techniques and professional judgment for the development of the base year system and for the derivation of variables that interface with the demand forecasts. In this process the occurrence of errors is inevitable. Second, model specification errors are found in the creation of the relationship between the demand and supply inputs to generate output demand forecasts. The misspecification may include the issues of study area determination, model segmentation, omission or incorrect design of key variables, future year changes in behavioral parameters, transferability, aggregation biases, and the scale factor problem. Third, external or exogenous errors are associated with the external inputs or assumptions that underpin the demand forecasting model. They include assumptions concerning external factors (e.g., gross domestic product or income growth), errors in the planning assumptions, ramp-up effects, definition of the do-minimum and do-something alternatives, and interactions with other transport operators and services (*5*).

This paper does not intend to address all the issues enumerated, but does focus on ramp-up effects. The issue of patronage ramp-up represents the delay in ridership take-up during the initial periods of a new transport service; the new service can include newly constructed and substantially upgraded services and infrastructure. In the ramp-up stage, the patronage increases relatively greatly overall while fluctuating within the period (*6*).

Ramp-up has not been incorporated in conventional travel demand models. Hence, there are difficulties in understanding the ramp-up itself, and as a result, the reliability of transport demand analyses has been compromised to that extent. This again has expanded the uncertainty of economic feasibility studies and has curtailed the confidence level of transport appraisal. Nevertheless, there are few studies in the literature to investigate that issue.

A novel ramp-up analysis model based on a heuristic *F*-test is developed in this paper. The model is applied to road and rail schemes of South Korea to examine ramp-up parameters. Some implications and suggestions of this study are put forward.

PATRONAGE RAMP-UP

Figure 1 shows a conceptual diagram for understanding the ramp-up effect. As stated in the introductory section, ramp-up refers to the delay of demand building during the initial start-up periods of a new transport service. Normally the new service represents the construction of new infrastructure, but any form of substantially upgraded transport supply can be included.

The ridership during the ramp-up stage shows a large increase, but the curve fluctuating rather than monotonic. These characteristics are contrasted with those of the steady state phase. In the steady state step, drastic patronage oscillations are unlikely to occur. There is also an argument that ramp-up is less aggressive than is assumed (*7*). The effect, however, is generally accepted as a key risk in traffic forecasting for transport projects (*8*).

There seems to be no firm consensus as to the exact causes of ramp-up effects, but three common reasons can be identified (*6*). First, the "learning curve" is the most frequently cited explanation for this effect. Travelers, in the process of utility maximization or cost minimization of transport choices, are understood to develop their final

J. S. Chang and K.-H. Jung, Graduate School of Environmental Studies, Seoul National University, 599 Gwanak-ro, Gwanak-gu, Seoul 151-742, South Korea. S.-B. Chung, Graduate School of Railroad, Seoul National University of Technology, 138 Gongneung-gil, Nowon-gu, Seoul 139-743, South Korea. K.-M. Kim, Korea Development Institute, Chongnyangri-dong, Dongdaemun-gu, P.O. Box 113, Seoul 130-012, South Korea. Corresponding author: J. S. Chang, jsc@snu.ac.kr.

Transportation Research Record: Journal of the Transportation Research Board, No. 2175, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 84–91.
DOI: 10.3141/2175-10

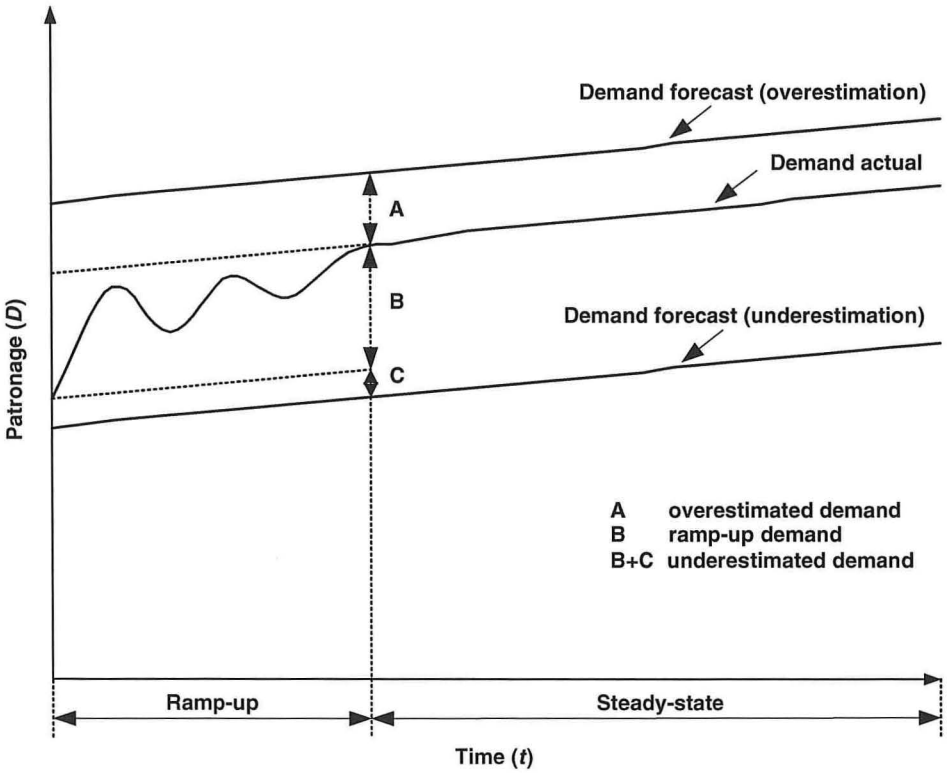


FIGURE 1 Conceptual diagram of patronage ramp-up.

decisions by learning from their mistakes. Second, people require some time for changing their travel behavior to respond to the opportunities afforded by a new travel option. Although this cause is not strictly separable from the learning curve effect, it seems to be more often used when demand shifting and induction are addressed. Finally, operational teething troubles are one of the typical reasons for the effect. Although the causes of learning curve and behavioral change are related to the viewpoint of trip makers, the last issue is associated with that of operators. The three reasons can be understood as the adaptation process of consumers and suppliers to new transport services. Thus, good marketing or service experimentation may reduce the ramp-up phenomenon.

There are also some external factors that affect the ramp-up effect: network influences, land use changes, and governmental policies. Network effects address existing or planned transport facilities, not including the proposed facility or service being evaluated, because those facilities can influence the quantity of demand shifts. Next, land use changes are more connected with induced demand because changes in land use vary the activities of travelers; the variation eventually would alter travel demand quantities and patterns. Finally, the effect of governmental policies is diverse. The schemes can be soft measures such as taxation and demand management or harder plans that involve additional network supply and land use changes.

Despite the general recognition of these factors, it is difficult to isolate which factors affect ramp-up for each project and to what degree. That task is not the purpose of this study. This paper concentrates on developing a ramp-up analysis model, of which the primary outputs are ramp-up parameters, including the period and the degree of the effects. Also, a conceptual bridge from this study to the cause-and-effect analysis is shown.

METHODOLOGY

Existing Approaches

There are not many existing frameworks for investigating ramp-up effects. Two methodologies, however, can be found from the literature: professional judgment and regression analysis.

The professional judgment approach examines patronage ramp-up on the basis of the insight and experience of experts. Thus, this approach could be classified as a form of Delphi survey: a group decision-making process based on the likelihood that a certain event will occur. The method makes use of a panel of experts. The responses of the group to a series of questionnaires are anonymous, and the group is provided with a summary of opinions before answering the next questionnaire. It is believed that the group will converge toward the best response through this consensus process, shown in Figure 2.

This approach has been adopted by a few ramp-up studies (*9, 10*). However, there is little theoretical background in this methodology and it is difficult to avoid arbitrary decisions.

Regression analyses are a more frequently used econometric technique than Delphi surveys (*6*). The approach selects a reference scheme that is already operational for a target transport supply. Next, the ridership data of the reference project are collected. After a regression line is drawn that is believed to best fit the observation, the duration and the ratio of ramp-up are determined (Figure 3). Finally, the two parameters are applied to the target scheme. However, the threshold dividing ramp-up and steady state steps is arbitrarily determined, as shown in Figure 3. The reason is that the econometric technique is based on a continuous function, and hence there are no internal mechanisms to find the splitting point.

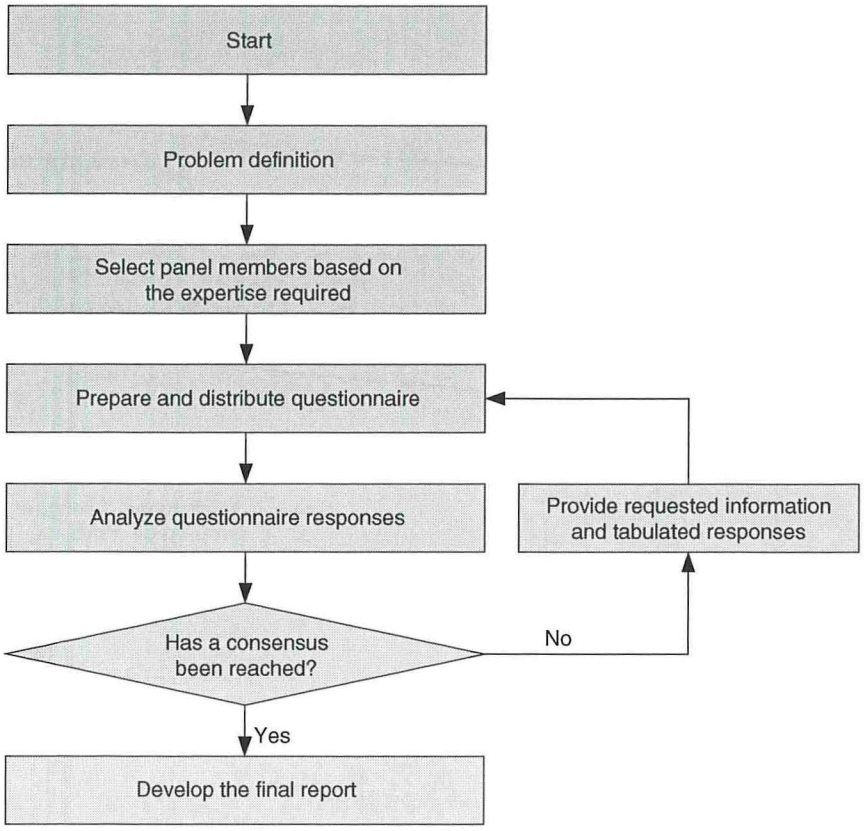


FIGURE 2 Flowchart for Delphi method.

Proposed Methodology

Let ridership D_t be the patronage using a particular transport service or infrastructure at time t where t is set as a sufficient period of time in which demand differences between observations appear. The ratio of patronage variation v_t is given as

$$v_t = \frac{D_t - D_{t-1}}{D_{t-1}} \quad (1)$$

The ratio is expected to show great oscillations during the first periods of a new service. This trend is also expected to gradually decrease as time goes by (Figure 4). Thus, it is not difficult to anticipate that the variances of increasing and decreasing rates between the ramp-up and steady state phases show statistical differences.

An F-test can be a useful tool to analyze the heteroscedasticity. This test is the most common statistical technique to deal with independent random samples from two populations. In this process, the incorporation of the upper and lower control limits is helpful to reduce biases from extreme values.

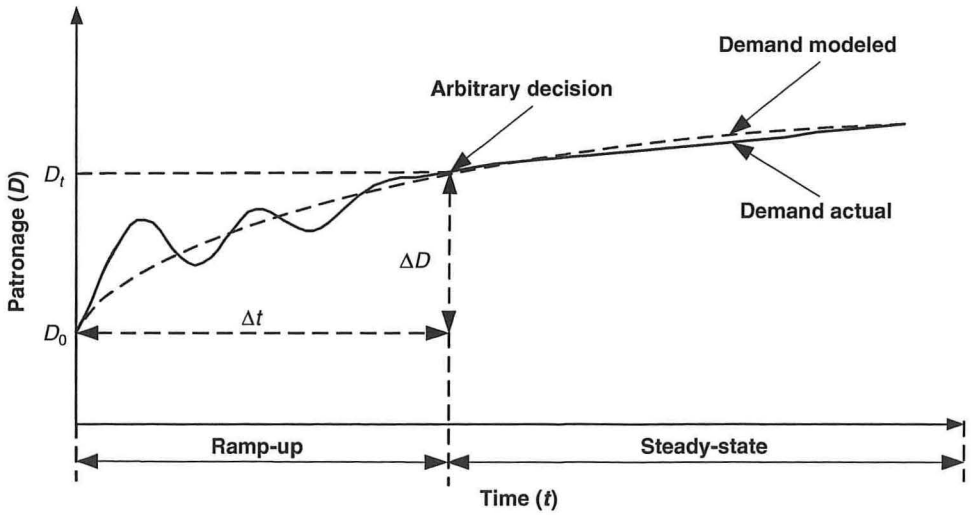


FIGURE 3 Conceptual diagram of regression analyses for ramp-up studies.

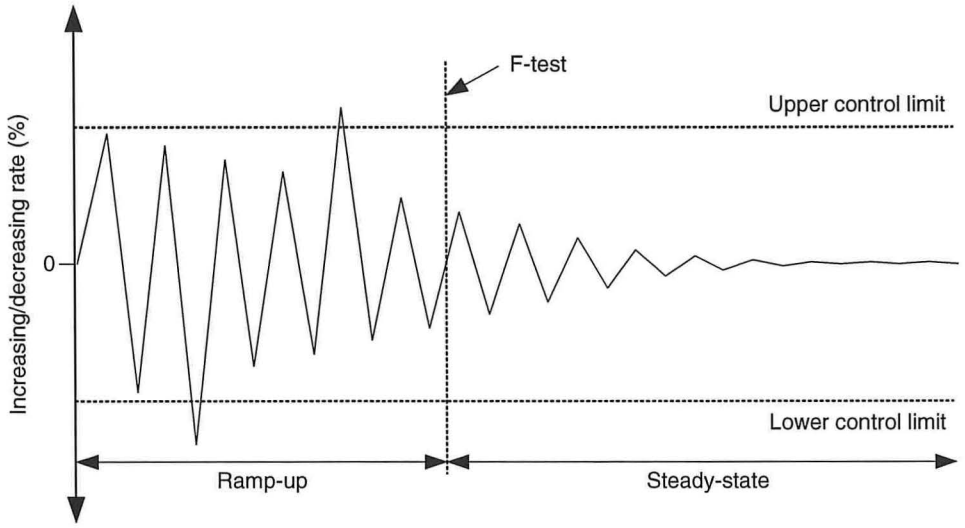


FIGURE 4 Conceptual diagram of *F*-test for ramp-up studies.

Let S_r^2 and S_s^2 be the variances of independent random samples of ramp-up and steady state stages of size n_r and n_s, respectively. Let σ_r^2 and σ_s^2 be the corresponding variances of populations. If S_r^2 and S_s^2 are taken from two normal populations having the same variance, then $F = S_r^2/S_s^2$ or S_s^2/S_r^2 is a random variable having the F distribution with parameters $\eta_r = n_r - 1$ and $\eta_s = n_s - 2$, where η_r and η_s are the numerator and denominator degrees of freedom.

The null hypothesis H_0 and the alternative hypothesis H_1 of the F-test for examining the heteroscedasticity can be given as follows:

$$H_0: \sigma_r^2 = \sigma_s^2$$
$$H_1: \sigma_r^2 > \sigma_s^2, \sigma_r^2 < \sigma_s^2, \text{or } \sigma_r^2 \neq \sigma_s^2 \quad (2)$$

Either a one-tailed or a two-tailed F-test can be used in this analysis. However, a one-sided criterion is a more convenient procedure in practice because F-values are always greater than one when the lower variance is positioned as the denominator. Thus, F-values are expected to be greater than one with the variance of the steady state phase as the denominator; the variance of the ramp-up stage is expected to be greater than that of the steady state phase, as shown in Figure 4. Hence, the alternative hypothesis in this study is set as H_1: $\sigma_r^2 > \sigma_s^2$ with a one-tailed test.

If the null hypothesis is rejected, namely, $F > F_\alpha$ where F_α represents F-values corresponding to right-hand tails of the level of significance α, which is commonly set as 0.05, then the result is understood as representing the ramp-up stage. Otherwise, if $F < F_\alpha$, then the steady state has been found through the test.

The issue, however, is to find the changing point of heteroscedasticity. A heuristic way is adopted in this paper (Figure 5). Thus, when the F-test at time t accepts the null hypothesis, the ramp-up period is determined until $t - 1$; otherwise the F-test is iterated with the change of time $t := t + 1$, assuming that the ramp-up continues.

Once the null hypothesis is accepted, the ramp-up ratio is defined as:

$$\Omega_t = \frac{D_t}{D_0 + \Delta D}$$

such that

$$0 \leq t \leq \Delta t \quad (3)$$

where Ω_t is the ramp-up ratio at time t, D_0 is the ridership at project opening, and ΔD is the increased patronage during the ramp-up period Δt.

CASE STUDIES

In this section the model suggested in the previous section is applied to road and rail schemes of South Korea. Specifically, the Cheonan–Nonsan (C-N) Motorway and the Seoul Metro Line 8 are selected.

Patronage take-up data are collected monthly. Seasonal variations of the data, however, should be removed; if yearly data are collected to avoid the variations, it is hard to guarantee the statistical significance of the developed model. Monthly adjustment factors, thus,

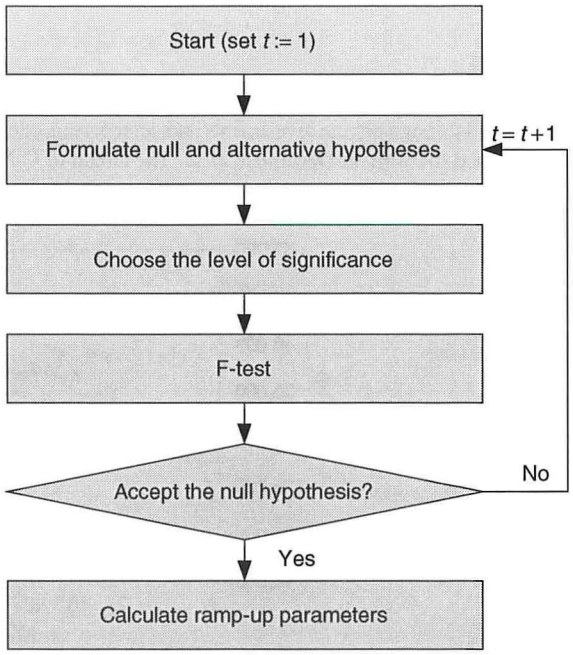


FIGURE 5 Flowchart for heuristic *F*-test.

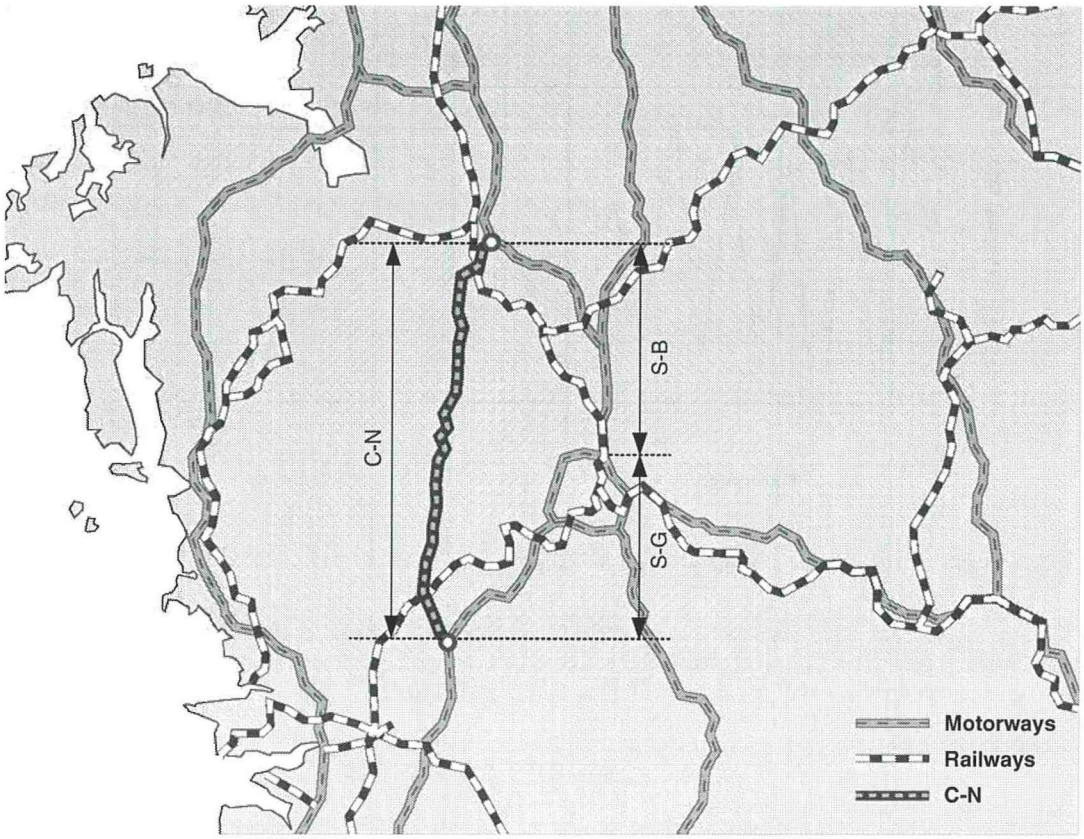


FIGURE 6 Map of C-N Motorway.

are used to reduce seasonal biases, which draw normalized demands. The factors for the road scheme are calculated from the nationwide traffic data of motorways, and the rail adjustment factors are arranged with the ridership data of the Seoul metropolitan electric railway system.

Road Scheme

C-N was funded through private investments and, after 5 years of construction, opened to traffic in December 2002. The length of C-N is 80.96 km and the posted speed is 100 km/h. The standard automobile toll charged on C-N is 96.73 won per kilometer. This is 2.4 times more than that of the Seoul–Busan (S-B) and Seoul–Gwangju (S-G) Motorways, which are both 40.2 won per kilometer. However, after the opening of the C-N, travelers can save 30 km compared with the previous route by using a combination of the S-B and S-G Motorways (Figure 6). The reduction is expected to save approximately half an hour of travel time.

Data sources for road traffic on the C-N are the monthly record of tollgate-to-tollgate volumes. Figure 7 shows the monthly average number of vehicles on C-N. Demand is steadily increasing after the

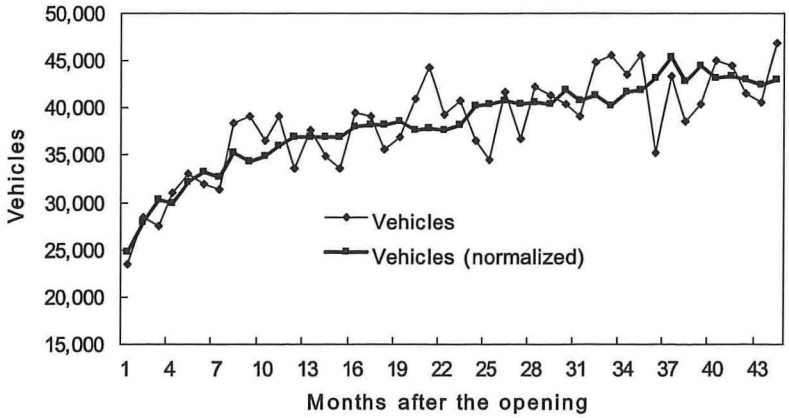


FIGURE 7 Monthly average numbers of vehicles on C-N.

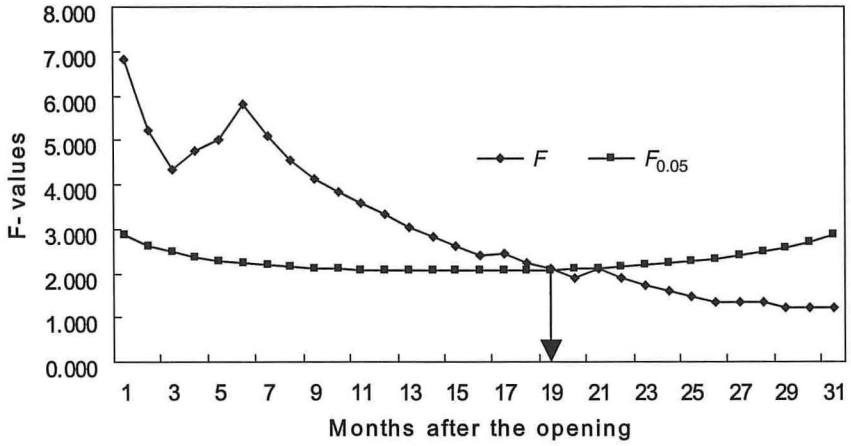


FIGURE 8 Result of heuristic *F*-test of C-N.

opening of the motorway. Seasonal variations are observed in some periods, but the fluctuations are smoothed by the adjustment factor, which yields the normalized demand.

The monthly variations were investigated by the heuristic *F*-test. The test found the dividing point between the ramp-up and steady state phases at 19 months after the initiation of the new road, as shown in Figure 8. On the basis of this result, the ramp-up ratio of C-N is calculated, as shown below:

Period (months)	*1*	*6*	*12*	*19*
Vehicles	24,797	33,238	37,056	36,979
(Ramp-up ratio)	0.67	0.89	0.99	1.00

FIGURE 9 Map of Line 8.

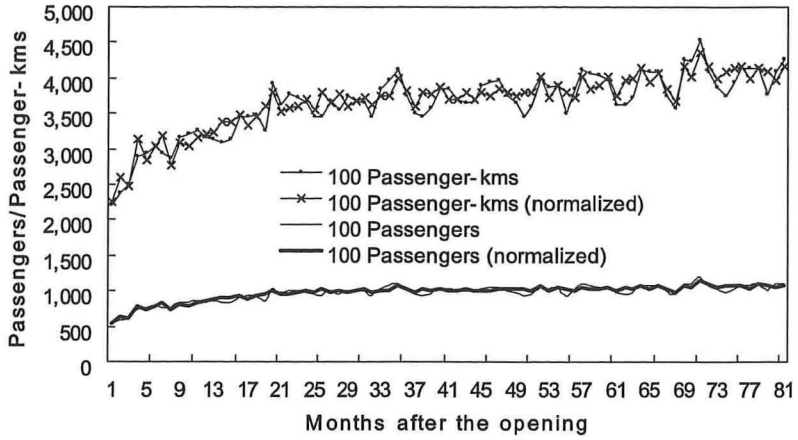


FIGURE 10 Monthly average number of trips on Line 8.

Rail Scheme

Line 8 belongs to the Seoul metropolitan electric railway system and was constructed in the southeast region of Seoul (Figure 9). The construction progressed in two phases. The first opened in November 1996 and involved stations from Moran to Jamsil. The second phase extended the line from Jamsil to Amsa, which opened in July 1999. In total, Line 8 has 17 stations spanning the 17.7-km operational length.

The ramp-up investigation in this paper deals with the demand variation after the initiation of the second phase. Thus, the examination physically covers 4.6 km of operational length from Jamsil to Amsa with five stations.

Data sources for rail travel by Line 8 are the monthly records of the rail use in the Seoul metropolitan area supplied by the Korea Railroad Corporation. Two typical types of data were used: passengers and passenger kilometers (Figure 10). Although the two indices of passengers and passenger kilometers display similar trends of patronage increases, a slightly larger oscillation is observed with the latter indicator. As with the case of the road example, the monthly adjustment factor smooths the seasonal variations.

Figure 11 shows the results of the heuristic *F*-test of the rail case. It indicates that ramp-up lasted approximately 2.5 years. Passengers and passenger kilometers display no substantial difference in the ramp-up parameter (Table 1). The cause could be inferred from the characteristics of Line 8 as an urban railway. With travel by urban mass transit, the journey distances for each trip have a relatively low variance. Thus, an analogous degree of ramp-up by the two indices is observed. Conversely, intercity rail travel should show larger differences in the ramp-up parameters between the two indicators.

DISCUSSION OF RESULTS

This research has suggested a novel ramp-up analysis model, but the nature of the effect has not been completely examined. The reason is that the developed model deals just with the parameters of ramp-up but does not involve the causal relationship of this effect. The functional relationship of the cause-and-effect analysis $f(\cdot)$ could be expressed as follows:

$$R(p,r) = f(X_c, X_b, X_o; Y_n, Y_l, Y_g) \tag{4}$$

where

$R(\cdot)$ = ramp-up effect with the parameters of period p and the ratio r;

X = the structural causes of ramp-up involving learning curve c, changes in travel behavior b, and operational teething problems o; and

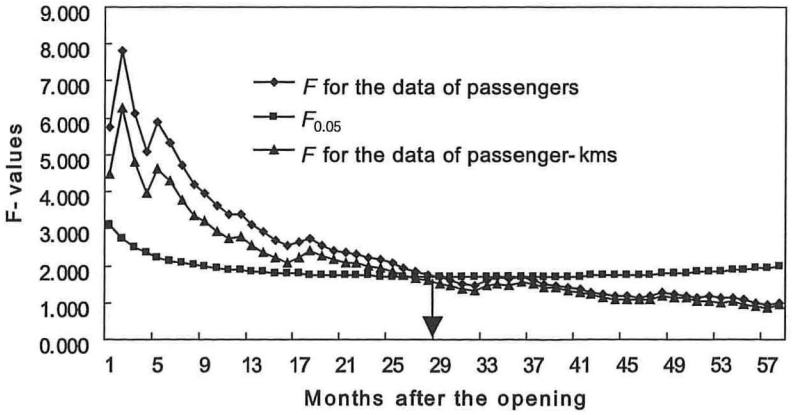


FIGURE 11 Result of the *F*-test of Line 8.

TABLE 1 Ramp-Up Ratio of Line 8

Period (months)	1	6	12	18	24	29	30
10^4 passengers (ramp-up ratio)	53,051 (0.52)	61,416 (0.60)	71,455 (0.70)	81,493 (0.80)	91,532 (0.90)	99,897 (0.98)	101,570 (1.00)
10^6 passenger kms (ramp-up ratio)	219,607 (0.59)	247,234 (0.66)	280,386 (0.75)	313,538 (0.84)	346,690 (0.93)	374,317 (1.00)	N/A

TABLE 2 Result of Simulation

Variable	Without Ramp-Up (million won)	With Ramp-Up (million won)
Travel time savings	89,052	87,558
Vehicle operating cost savings	203,032	194,932
Traffic accident cost savings	30,742	29,882
Environmental cost savings	18,027	17,612
Total benefits discounted	104,172	96,952
Total costs discounted	102,688	102,688
Benefit–cost ratio	1.01	0.94

NOTE: 1,000 won ≈ 1 US$.

Y = external causes of ramp-up including network effects n, land-use changes l, and governmental policies g.

Through the results of this study, the two dependent variables, p and r, of a new transport scheme could be estimated. This means that the heuristic F-test for a scheme can generate only one observation for the causal investigation. This also means that sufficient numbers of schemes should be examined to have necessary data points for the causality study; note that collection of reliable data for the independent variables is another burden. Further adding to the difficulty is that there is no established consensus on the explicit functional relationship that can reasonably substitute the indirect functional form. Thus, the causal relationship of ramp-up requires another dimension of research that is set as a future study.

Nevertheless, it would be useful to apply the ramp-up effect to an economic feasibility study. A simple numerical example is given. The simulation assumes that ramp-up lasts 4 years and each year is designated its own ramp-up ratio; the opening year 50%, the second year 70%, the third year 90%, and the fourth year 100% ramp-up proportions, respectively.

Table 2 shows the simulation results of economic appraisal for a scheme with or without consideration of the ramp-up effect. The experiment shows that total benefits of the "with ramp-up" case decrease approximately 7% with no changes in the cost. This results in a change in the benefit–cost ratio from 1.01 to 0.94. In general, everything else being controlled, the decision-making threshold of economic feasibility is the B:C ratio 1.0. The experiment addresses that the incorporation of patronage ramp-up would supply more cautious decision making for investments.

CONCLUSION

Ramp-up represents the delay in ridership take-up during the first periods of the opening of a new transport service or facility. In this stage, ridership shows larger oscillations compared with that of the steady state phase. Traditionally, professional judgments and regression analyses have been used to examine this effect. The methodologies, however, have inherent limitations of being biased and arbitrary. As an alternative, this study has proposed a heuristic F-test that finds the changing point of heteroscedasticity with statistical reliability between the ramp-up and steady state stages. The framework suggested has been successfully applied to road and rail schemes of South Korea. The ramp-up parameters consisting of the duration and the ratio of this effect have also been calculated.

It is hoped that the model will be employed as a useful tool for ramp-up studies. However, as discussed previously, the nature of patronage ramp-up has not been completely examined yet. The reason is mainly that the causal relationship of this effect could not be included in this study and, hence, follow-up studies are required. This continuous research offers more opportunities for addressing and applying ramp-up effects to transport studies. This, of course, would reduce the error of transport demand forecasts and could increase the reliability of economic feasibility tests.

REFERENCES

1. Mackie, P., and J. Preston. Twenty-One Sources of Error and Bias in Transport Project Appraisal. *Transport Policy,* Vol. 5, 1998, pp. 1–7.
2. Flyvbjerg, B. Measuring Inaccuracy in Travel Demand Forecasting: Methodological Considerations Regarding Ramp Up and Sampling. *Transportation Research Part A,* Vol. 39, No. 6, 2005, pp. 522–530.
3. Flyvbjerg, B., M. Holm, and S. Buhl. How (In)accurate Are Demand Forecasts in Public Works Projects? *Journal of the American Planning Association,* Vol. 71, 2005, pp. 131–146.
4. Flyvbjerg, B., M. Holm, and S. Buhl. Inaccuracy in Traffic Forecasts. *Transport Reviews,* Vol. 26, 2006, pp. 1–24.
5. *Demand Forecasting Errors.* Transport Note No. TRN-12. World Bank, Washington, D.C., 2005.
6. Douglas, N. *Patronage Ramp-Up Factors for New Rail Services.* Working paper. Douglas Economic, Ltd., 2003.
7. Bain, R., and L. Polakovic. *Traffic Forecasting Risk Study Update 2005: Through Ramp-Up and Beyond.* Standard and Poor's Research Report—Corporates, New York, 2005.
8. Kriger, D., S. Shiu, and S. Naylor. *NCHRP Synthesis of Highway Practice 364: Estimating Toll Road Demand and Revenue.* Transportation Research Board of the National Academies, Washington, D.C., 2006.
9. Howard, Needles, Tammen and Bergendoff Corporation and Transport Economics and Management Systems, Inc. *Orlando–Miami Planning Study.* Florida High Speed Rail Authority, Tallahassee, Fla., 2003.
10. Douglas Economics (formerly PCIE). *East Coast Corridor High Speed Rail Patronage Forecasts.* Commonwealth Department of Transport and Regional Services, Sydney, Australia, New Zealand, 2002.

The Transportation Demand Forecasting Committee peer-reviewed this paper.

Analysis of Implicit Choice Set Generation Using a Constrained Multinomial Logit Model

Michel Bierlaire, Ricardo Hurtubia, and Gunnar Flötteröd

Discrete choice models are defined conditional to the analyst's knowledge of the actual choice set. The common practice for many years has been to assume that individual-based choice sets can be deterministically generated on the basis of the choice context and characteristics of the decision maker. This assumption is not valid or not applicable in many situations, and probabilistic choice set formation procedures must be considered. The constrained multinomial logit model (CMNL) has recently been proposed as a convenient way to deal with this issue, as it is also appropriate for models with a large choice set. In this paper, how well the implicit choice set generation of the CMNL approximates the explicit choice set generation is analyzed as described in earlier research. The results based on synthetic data show that the implicit choice set generation model may be a poor approximation of the explicit model.

In standard choice models, it is assumed that the alternatives considered by the decision maker can be deterministically specified by the analyst. The choice set is characterized by deterministic rules based on the characteristics of the decision maker and the choice context. For example, single-room apartments are not considered by families with children in a house choice context, and a car is not considered as a possible transportation mode if the traveler has no driver's license or no car.

There are, however, many situations in which the deterministic choice set generation procedure is not satisfactory, or even possible. Data may be unavailable (the number of children in the household is unknown to the analyst), or rules are fuzzy by nature. For instance, train is not considered as a transportation mode if it involves a long walk to reach the train station. But how long is a "long walk"?

Modeling explicitly the choice set generation process involves a combinatorial complexity, which makes the models intractable except for some specific instances. Manski defines the theoretical framework in a two stage process (*1*), in which the probability that decision maker n chooses alternative i is given by

$$P_n(i) = \sum_{C_m \subseteq C} P_n(i|C_m) P_n(C_m) \tag{1}$$

Transport and Mobility Laboratory, School of Architecture, Civil and Environmental Engineering, Ecole Polytechnique Fédérale de Lausanne, Lausanne 1015, Switzerland. Corresponding author: M. Bierlaire, michel.bierlaire@epfl.ch.

Transportation Research Record: Journal of the Transportation Research Board, No. 2175, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 92–97.
DOI: 10.3141/2175-11

where $P_n(i|C_m)$ is the probability for individual n to choose alternative i conditional to the choice set C_m, and $P_n(C_m)$ is the probability for individual n to consider choice set C_m. The sum runs on every possible subset C_m of the universal choice set C.

Swait and Ben-Akiva (*2*) and Ben-Akiva and Boccara (*3*) build on this framework and use explicit random constraints to determine the choice set generation probability. The probability of considering a choice set C_m is a function of the consideration of the different alternatives in the universal choice set as follows:

$$P_n(C_m) = \frac{\prod_{i \in C_m} \phi_{in} \prod_{j \in C_m} (1 - \phi_{jn})}{1 - \prod_{k \in C} (1 - \phi_{kn})} \tag{2}$$

where ϕ_{in} is the probability that alternative i is considered by user n, which may be modeled by a binary logit model that depends on the alternative's attributes. Note that Equation 2 assumes independence of the consideration probabilities across alternatives, which is a restrictive assumption because there can be correlation in the consideration criteria of different alternatives.

Swait proposes to model the choice set generation as an implicit part of the choice process in a multivariate extreme value framework, requiring no exogenous information (*4*). Here, choice sets are not separate constructs but another expression of preferences. The probability of considering a choice set is defined as the probability for that choice set to correspond to the maximum expected utility for an individual n:

$$P_n(C_m) = \frac{e^{\mu I_{n,C_m}}}{\sum_{C_k \subseteq C} e^{\mu I_{n,C_k}}} \tag{3}$$

where μ is the scale parameter for the higher level decision (choice set selection) and I_{n,C_m} is the inclusive value (the "logsum" or expected maximum utility) of choice set C_m for decision maker n:

$$I_{n,C_m} = \frac{1}{\mu_m} \ln \sum_{j \in C_m} e^{\mu_m V_{nj}} \tag{4}$$

Here μ_m is the scale parameter and V_{nj} is the deterministic utility of alternative j for decision maker n. Swait's probabilistic choice set generation approach does not require assumptions by the analyst about which attributes affect an alternative's availability. Note that Swait's model also needs to account for every possible subset C_m of the universal choice set C.

Clearly, these methods are hardly applicable to medium- to large-scale choice problems because of the computational complexity that

arises from the combinatorial number of possible choice sets. If the number of alternatives in the universal choice set is J, the number of possible choice sets is $(2^J - 1)$.

In the context of route choice, Frejinger et al. assume that all decision makers consider the universal choice set, so that $P_n(C_m) = 0$ when $C_m \neq C$, and only one term remains in Equation 1 (*5*). However, this may not be appropriate in other contexts.

Therefore, various heuristics have been proposed in the literature that derive tractable models by approximating the choice set generation process.

In the quantitative marketing literature, the use of heuristics to model the construction of the choice set (or consideration set) has been a usual practice; a review of existing models can be found in Hauser et al. (*6*). Many heuristics are based on lexicographic preferences rules [Dieckmann et al. (*7*)] in which the choice set is determined by key attributes of the alternatives on which consumers base the construction of their consideration set. This approach is similar to the elimination by aspects heuristic, proposed by Tversky (*8*). Models like the one proposed by Gilbride and Allenby consider the construction of the choice set as a two-stage process (*9*), which is consistent with Manski's approach but solves the choice set enumeration issue by using Bayesian and Monte Carlo estimation methods.

Other heuristics use a one-stage approach [see, for example, Elrod et al. (*10*)] in which the choice set generation process is simulated through direct alternative elimination. This is done by setting the alternative's utility to minus infinity when certain attributes reach a threshold value. The alternative-elimination approach implies a different behavioral assumption from the two-stage approach, in which the individual does not observe choice sets explicitly but, instead, makes a compensatory choice between all alternatives belonging to a unique choice set of available or "possible" alternatives, which is a subset of the universal choice set.

Following the same one-stage approach, other heuristics assume that the elimination of alternatives is not deterministic. These are based on the use of penalties in the utility functions and have been proposed by Cascetta and Papola (*11*) [the implicit availability and perception (IAP) model] and expanded by Martinez et al. (*12*) [the constrained multinomial logit (CMNL) model]. In the next section, the CMNL model is briefly described and its theoretical background in the context of choice set generation is provided. Next the CMNL is compared with the theoretical framework (Equation 1), first through a simple example and, second, by estimating both models on synthetic data. The paper ends with conclusions and a discussion of further work.

CHOICE SET GENERATION WITH CMNL MODEL

Assuming that C_n is the choice set that the decision maker is actually considering, the choice model is given by

$$P_n(i|C_n) = \Pr(U_{in} \geq U_{jn} \qquad \forall j \in C_n) \tag{5}$$

where U_{in} is the random utility associated with alternative i by decision maker n. If C_n is known to the analyst, it can be characterized by indicators of the consideration of each alternative by the decision maker:

$$A_{in} = \begin{cases} 1 & \text{if alternative } i \text{ is considered by individual } n \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

The choice model can be equivalently written as

$$\begin{aligned} P_n(i|C_n) &= \Pr(U_{in} \geq U_{jn} \qquad \forall j \in C_n) \\ &= \Pr(U_{in} + \ln A_{in} \geq U_{jn} + \ln A_{jn} \qquad \forall j \in C) \end{aligned} \tag{7}$$

For an unconsidered alternative, this adds $\ln 0 = -\infty$ to its utility, so that the choice probability is 0, whereas the addition of $\ln 1 = 0$ has no effect on the utility of a considered alternative.

In the case of a logit model, the choice probabilities are

$$P_n(i) = \frac{e^{V_{in} + \ln A_{in}}}{\sum_{j \in C} e^{V_{jn} + \ln A_{jn}}} \tag{8}$$

The heuristics proposed by Cascetta and Papola (*11*) and Martinez et al. (*12*) consist in replacing the indicators A_{in} by the probability ϕ_{in} that individual n considers alternative i.

Cascetta and Papola introduce the IAP model as a way to incorporate awareness of paths into route choice modeling without requiring an explicit choice set generation step (*11*). A similar approach that penalizes the utilities of "dominated" alternatives is proposed by Cascetta et al. (*13*).

Martinez et al. expand the IAP idea and propose the CMNL model (*12*). The functional form for ϕ_{in} is assumed to be a binary logit, considering that the availability of an alternative is related with bound constraints on its attributes. For example, if X_{ink} is the kth variable of alternative i for decision maker n that influences the consideration of i, one has

$$\phi^u_{in}(X_{ink}; u_k, \omega_k) = \frac{1}{1 + \exp(\omega_k (X_{ink} - u_k))} \tag{9}$$

where the u_k parameter is the value at which the constraint is most likely to bind, and ω_k is the scale parameter of the binary logit. For instance, X_{ink} may be the walking distance to the train station, and u_k may be the maximum distance that individual n is willing to walk. Both u_k and ω_k are to be estimated. The intuition is that when the attribute X_{ink} exceeds u_k, the consideration probability ϕ^u_{in} tends to zero, while this availability tends to one when the value of the attribute is below u_k.

Equation 9 represents an upper value cutoff, where u_k represents the maximum value that the attribute X_{ink} can have for alternative i to be considered. To model a lower value cutoff, one needs only to invert the sign of the scale parameter ω_k:

$$\phi^l_{in}(X_{ink}; \ell_k, \omega_k) = \frac{1}{1 + \exp(-\omega_k (X_{ink} - \ell_k))} \tag{10}$$

where ℓ_k is the lower bound, which is analogous to u_k (upper bound) in Equation 9.

Functions 9 and 10 can be generalized to account for more than one constraint, allowing for several upper and lower bounds to be included simultaneously:

$$\phi_{in}(X_{in}; \ell, u, \omega) = \prod_k \phi^u_{in}(X_{ink}; u_k, \omega_k) \phi^l_{in}(X_{ink}; \ell_k, \omega_k) \tag{11}$$

The CMNL approach has an operational advantage over Manski's framework because it does not require enumerating the choice sets, which makes it easier to specify and estimate. However, the CMNL model is a heuristic that is based on convenient assumptions about

the functional form of the utility function. That is why the CMNL model can at most be considered as an approximation of Manski's model. The next section evaluates the quality of this approximation.

COMPARISON OF CMNL WITH MANSKI'S MODEL

This section compares the CMNL model with Manski's model. For this, a simple example is first presented in which the difference between the choice probabilities obtained by using both models is analyzed. Second, the CMNL model and Manski's model are estimated over synthetic data and the results are compared. For notational simplicity, the index n is subsequently omitted for the decision maker.

Simple Example

Consider a logit model with only two alternatives, where Alternative 1 is always considered ($\phi_1 = 1$) and Alternative 2 has probability ϕ_2 of being considered by the decision maker. Figure 1 shows the structure of Manski's framework if every possible combination of alternatives is considered as a choice set. This simple situation corresponds to a case in which the decision maker is captive to Alternative 1 with probability $1 - \phi_2$ [see also the captivity logit model proposed by Gaudry and Dagenais (*14*)].

The CMNL model defines the probability of choosing Alternative 1 as

$$P(1) = \frac{e^{V_1}}{e^{V_1} + e^{V_2 + \ln \phi_2}} \tag{12}$$

Manski's model (Equation 1) defines the probability of choosing Alternative 1 as

$$P(1) = P(\{1\})\frac{e^{V_1}}{e^{V_1}} + P(\{1,2\})\frac{e^{V_1}}{e^{V_1} + e^{V_2}} \tag{13}$$

where $P(\{1\})$ is the probability of considering the choice set composed only of Alternative 1 and $P(\{1,2\})$ is the probability of considering the choice set containing both alternatives. According to Equation 2, the choice set probabilities are

$$P(\{1\}) = \frac{\phi_1(1-\phi_2)}{1-(1-\phi_1)(1-\phi_2)} = 1 - \phi_2 \tag{14}$$

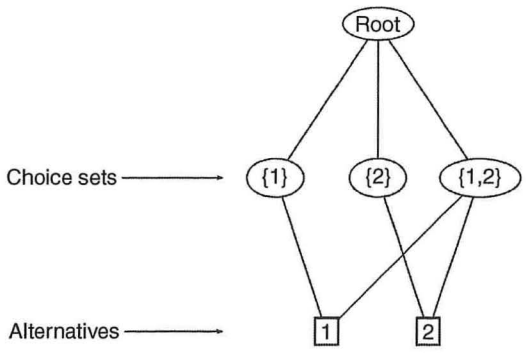


FIGURE 1 Example of a model in Manski's framework.

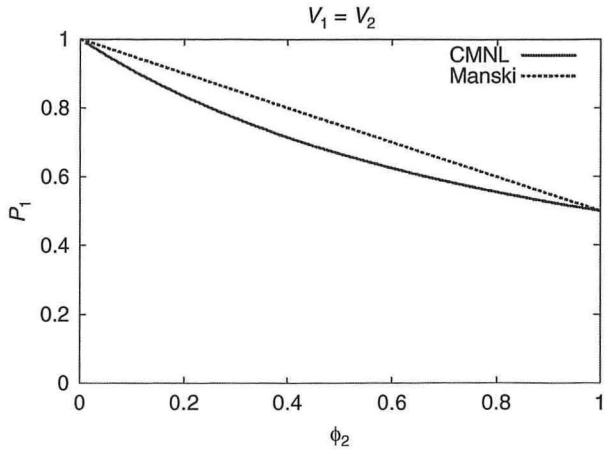


FIGURE 2 Choice probability of Alternative 1 (V1 = V2).

and

$$P(\{1,2\}) = \frac{\phi_1\phi_2}{1-(1-\phi_1)(1-\phi_2)} = \phi_2 \tag{15}$$

The probability of considering choice set {2} is 0 because Alternative 1 is always available. Therefore, Equation 13 becomes

$$P(1) = (1-\phi_2) + \phi_2\frac{e^{V_1}}{e^{V_1} + e^{V_2}} \tag{16}$$

In the deterministic limit ($\phi_2 = 0$ or $\phi_2 = 1$), the two models are equivalent. However, this is not the case anymore when ϕ_2 takes values between zero and one. The resulting choice probabilities are shown in Figure 2, assuming the same utility level V1 = V2 for both alternatives.

This figure shows that the CMNL is a good approximation of Manski's model only when φ2 is close to either zero or one, but it underestimates the probability of Alternative 1 elsewhere. If the utility for Alternative 1 is larger than the utility for Alternative 2 (Figure 3), the approximation improves. This makes sense because the more an alternative is dominated, the less important it is to know whether it really belongs to the choice set.

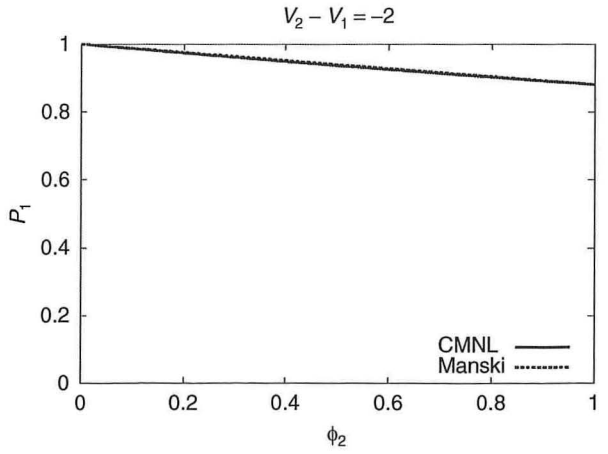


FIGURE 3 Probability of Alternative 1 (V1 > V2).

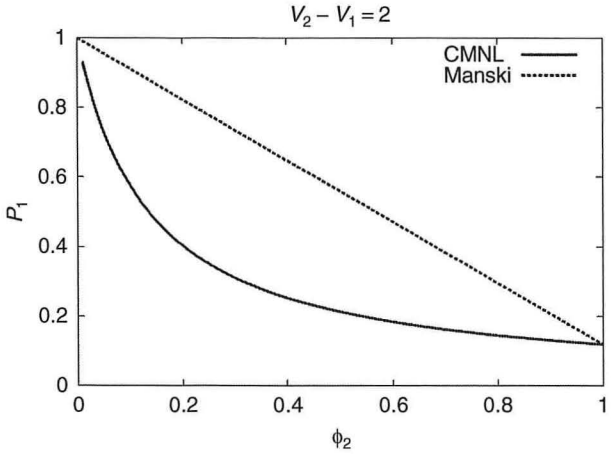


FIGURE 4 Choice probability of Alternative 1 (V1 < V2): V2 − V1 = 2.

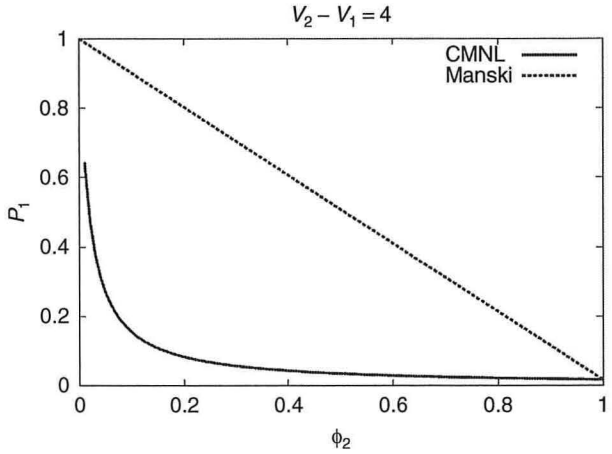


FIGURE 5 Choice probability of Alternative 1 (V1 < V2):V2 − V1 = 4.

However, as the utility of Alternative 1 becomes smaller and smaller compared with the utility of Alternative 2, the CMNL becomes a poorer and poorer approximation of Manski's model for intermediate ϕ_2 values, which is demonstrated in Figures 4 and 5.

These results can be interpreted as an unwanted compensatory effect in the CMNL model. The availability constraint is enforced by modifying the utility of the constrained alternative. However, when the utility of this alternative is high, it compensates the penalty. This means that the use of the CMNL model as an efficient choice set generation mechanism requires the assumption that the consideration probability for an alternative grows with its utility, meaning that the choice set depends only on the preferences of the individual. But alternatives with a high utility may be discarded in the presence of constraints such as budget or physical constraints. In the context of repetitive choices during a long period the individual may try to change her initial constraints to make the high-utility alternative available (for example, if the train produces high utility, a user may consider moving his residence closer to the train station), but in an instantaneous or short-term decision this may not be possible. That motivates one to analyze the performance of the CMNL on synthetic data, which is shown in the next section.

Synthetic Data

This section describes a series of controlled experiments in which some of the data are synthetically generated. Beginning with a real stated preference data set that was collected for the analysis of a hypothetical high-speed train in Switzerland (*15*), the alternatives are

1. Car (CAR),
2. Regular train (TRAIN), and
3. Swissmetro (SM), the future high-speed train.

From this data set, which consists of 5,607 observations, the attributes of the alternatives are used and synthetic choices are simulated on the basis of a postulated "true" model: a logit model with linear-in-parameters utility functions. The specification table as well as the "true" values of the parameters are reported in Table 1. The values have been obtained by estimating the model with real choices and by rounding the estimates.

It is assumed that the TRAIN and the SM alternatives are always considered, whereas the consideration of the CAR alternative depends on the travel time according to:

$$\phi_{\text{CAR}} = \frac{1}{1+\exp\left(\omega\left(\frac{\text{TT}_{\text{CAR}}}{60-a}\right)\right)} \tag{17}$$

which states that the probability of considering CAR as an available alternative decreases with the travel time TT_{CAR}, in minutes, and that this probability is .5 when the availability threshold a, in hours, is reached.

TABLE 1 Parameter Descriptions and Values

Parameter	Value	Car	Train	Swissmetro
ASC_{CAR}	0.3	1	0	0
ASC_{SM}	0.4	0	0	1
β_{cost}	−0.01	Cost (CHF)	Cost (CHF)	Cost (CHF)
β_{time}	−0.01	TT_{CAR} (min)	TT_{CAR} (min)	TT_{CAR} (min)
β_{he}	−0.005	0	Headway (min)	Headway (min)
α	3	Consideration threshold of car (h)		
ω	1,2,3,5,10	Consideration dispersion of car		

NOTE: ASC = alternative specific constant; CHF = Swiss franc.

This implies that, depending on the availability of the CAR alternative, there are two possible choice sets: the full choice set and the choice set containing only the TRAIN and the SM alternative. The random constraints approach [as stated in Ben-Akiva and Boccara (*3*)] defines the probability of each choice set as follows:

$$P(\{\text{TRAIN, SM}\}) = \frac{\phi_{\text{TRAIN}}\phi_{\text{SM}}(1-\phi_{\text{CAR}})}{1-(1-\phi_{\text{CAR}})(1-\phi_{\text{TRAIN}})(1-\phi_{\text{SM}})} = 1-\phi_{\text{CAR}} \quad (18)$$

and accordingly

$$P(\{\text{CAR, TRAIN, SM}\}) = \phi_{\text{CAR}} \quad (19)$$

The synthetic choices are generated by (*a*) simulating a choice set for each decision maker according to Equations 18 and 19 and (*b*) simulating a choice set for each decision maker by using the "true" model specified in Table 1.

One hundred choice data sets are simulated for each value of ω. These values generate constraints with different levels of uncertainty. Figure 6 shows the shape of these constraint functions. Estimation results for the Manski and the CMNL models are given in Tables 2 and 3. For each parameter β, the average value $\bar{\beta}$ and the standard error σ over 100 simulations are computed. In the tables, both $\bar{\beta}$ and the *t*-statistic $(\bar{\beta}-\beta)/\sigma$ are reported, the latter value being used to test whether the estimated value is significantly different from the true one. Because the tested hypothesis is that the average estimated value is equal to the "true" one, a low value of the *t*-statistic indicates that the estimate is not significantly different from the real parameter.

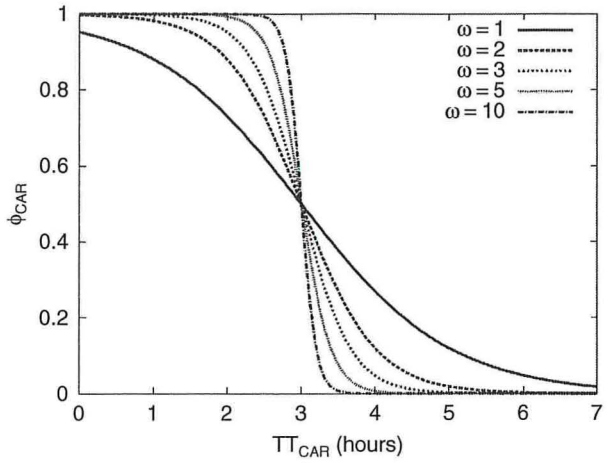


FIGURE 6 Shape of constraint for values of ω.

The estimates of Manski's model are unbiased. The hypothesis that the true value of any parameter is equal to the postulated value at the 95% level cannot be rejected. Several estimates of the CMNL model are biased (marked with *); the hypothesis that the true value of the parameter is equal to the postulated value being rejected at the 95% level. The quality of the CMNL estimates improves with decreasing dispersion (increasing ω). This is consistent with the findings in the section describing the simple example.

Figures 7 and 8 show the *t*-statistics for the cost and travel time parameter over different ω values for Manski's model and the

TABLE 2 Estimation Results for Manski's Model

Real ω Value		1		2		3		5		10	
Parameter	Real Value	Estimate	*t*-Test	Estimate	*t*-Test	Estimate	*t*-Test	Estimate	*t*-Test	Estimate	*t*-Test
ASC_{CAR}	0.3	0.304	0.027	0.288	0.113	0.300	0.010	0.301	0.012	0.314	0.184
ASC_{SM}	0.4	0.396	0.044	0.399	0.010	0.405	0.053	0.401	0.017	0.410	0.151
β_{cost}	−0.01	−0.010	0.283	−0.010	0.001	−0.010	0.179	−0.010	0.052	−0.010	0.012
β_{he}	−0.005	−0.005	0.241	−0.005	0.01	−0.005	0.048	−0.005	0.082	−0.005	0.078
β_{time}	−0.01	−0.010	0.074	−0.010	0.05	−0.010	0.049	−0.010	0.003	−0.010	0.001
α	3	2.963	0.019	3.008	0.118	3.000	0.100	2.998	0.081	3.002	0.101
ω	See top row	1.003	0.028	2.014	0.079	3.066	0.210	5.095	0.170	10.52	0.353

TABLE 3 Estimation Results for CMNL Model

Real ω Value		1		2		3		5		10	
Parameter	Real Value	Estimate	*t*-Test	Estimate	*t*-Test	Estimate	*t*-Test	Estimate	*t*-Test	Estimate	*t*-Test
ASC_{CAR}	0.3	0.503	0.950	0.421	1.153	0.406	1.365	0.380	0.988	0.326	0.313
ASC_{SM}	0.4	0.565	2.013[a]	0.550	2.375[a]	0.536	1.804	0.506	1.485	0.463	0.872
β_{cost}	−0.01	−0.008	4.825[a]	−0.008	3.580[a]	−0.009	2.309[a]	−0.009	1.182	−0.010	0.613
β_{he}	−0.005	−0.005	0.202	−0.005	0.151	−0.005	0.071	−0.005	0.120	−0.005	0.090
β_{time}	−0.01	−0.007	3.929[a]	−0.008	3.645[a]	−0.008	2.813[a]	−0.009	2.316[a]	−0.009	1.523
α	3	2.186	1.753	2.656	3.073[a]	2.773	3.762[a]	−2.869	3.305[a]	2.948	1.864
ω	See top row	1.043	0.239	2.094	0.403	3.118	0.431	5.238	0.424	12.146	3.149[a]

[a]Indicates a biased parameter.

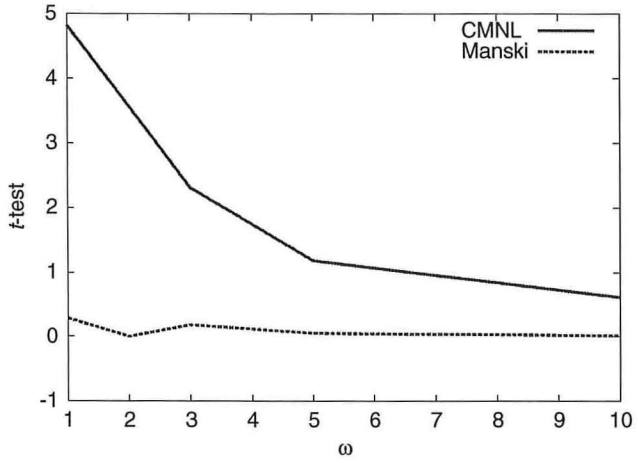


FIGURE 7 *t*-statistics for cost parameter over ω.

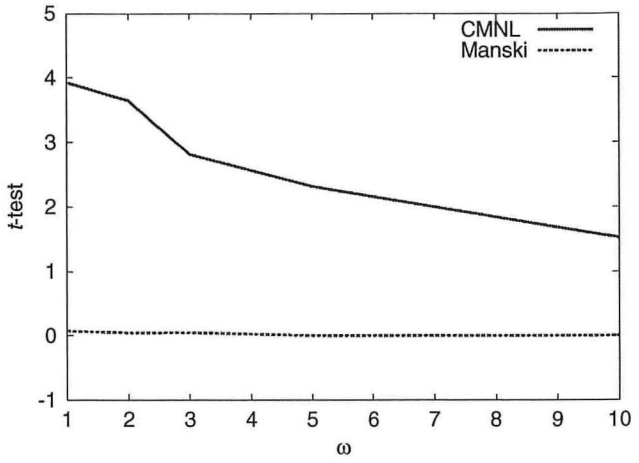


FIGURE 8 *t*-statistics for time parameter over ω.

CMNL model. The quality of the estimates is constant across different values of ω for Manski's model. The quality of the CMNL estimates increases with ω, and their *t*-statistics reach acceptable values when the constraint function becomes steep.

CONCLUSIONS AND FURTHER WORK

It has been shown with simple examples that the CMNL model is not adequate to model the choice set generation process consistently with Manski's framework. Consequently, the CMNL model should be considered as a model on its own, derived from semicompensatory assumptions as described by Martinez et al. (*12*), but not as a way to capture the choice set generation process. Its complexity is linear with the number of alternatives, whereas Manski's framework exhibits an exponential complexity.

An investigation has begun to determine whether a modified version of the CMNL could better approximate Manski's framework, but it has been unsuccessful so far. The derivation of a good approximation of Manski's model with the complexity of the CMNL would be particularly useful to handle models with a large number of alternatives.

REFERENCES

1. Manski, C. The Structure of Random Utility Models. *Theory and Decision,* Vol. 8, 1977, pp. 229–254.
2. Swait, J., and M. Ben-Akiva. Incorporating Random Constraints in Discrete Models of Choice Set Generation. *Transportation Research Part B,* Vol. 21, No. 2, 1987, pp. 91–102.
3. Ben-Akiva, M. E., and B. Boccara. Discrete Choice Models with Latent Choice Sets. *International Journal of Research in Marketing,* Vol. 12, 1995, pp. 9–24.
4. Swait, J. Choice Set Generation Within the Generalized Extreme Value Family of Discrete Choice Models. *Transportation Research Part B,* Vol. 35, No. 7, 2001, pp. 643–666.
5. Frejinger, E., M. Bierlaire, and M. Ben-Akiva. Sampling of Alternatives for Route Choice Modeling. *Transportation Research Part B,* Vol. 43, No. 10, 2009, pp. 984–994.
6. Hauser, J., M. Ding, and S. Gaskin. Non-Compensatory (and Compensatory) Models of Consideration-Set Decisions. *Proc., Sawtooth Software Conference,* Delray Beach, Fla., 2009.
7. Dieckmann, A., K. Dippold, and H. Dietrich. Compensatory Versus Noncompensatory Models for Predicting Consumer Preferences. *Judgment and Decision Making,* Vol. 4, 2009, pp. 200–213.
8. Tversky, A. Elimination by Aspects: A Theory of Choice. *Psychological Review,* Vol. 79, 1972, pp. 281–299.
9. Gilbride, T. J., and G. M. Allenby. A Choice Model with Conjunctive, Disjunctive, and Compensatory Screening Rules. *Marketing Science,* Vol. 23, No. 3, 2004, pp. 391–406.
10. Elrod, T., R. D. Johnson, and J. White. A New Integrated Model of Noncompensatory and Compensatory Decision Strategies. *Organizational Behavior and Human Decision Processes,* Vol. 95, No. 1, 2004, pp. 1–19.
11. Cascetta, E., and A. Papola. Random Utility Models with Implicit Availability/Perception of Choice Alternatives for the Simulation of Travel Demand. *Transportation Research Part C,* Vol. 9, No. 4, 2001, pp. 249–263.
12. Martinez, F., F. Aguila, and R. Hurtubia. The Constrained Multinomial Logit Model: A Semi-Compensatory Choice Model. *Transportation Research Part B,* Vol. 43, No. 3, 2009, pp. 365–377.
13. Cascetta, E., F. Pagliara, and K. Axhausen. The Use of Dominance Variables in Choice Set Generation. *Proc., 11th World Conference on Transport Research,* University of California at Berkeley, 2007.
14. Gaudry, M., and M. Dagenais. The Dogit Model. *Transportation Research Part B,* Vol. 13, No. 2, 1979, pp. 105–111.
15. Bierlaire, M., K. Axhausen, and G. Abay. Acceptance of Modal Innovation: The Case of the SwissMetro. *Proc., 1st Swiss Transportation Research Conference,* Ascona, Switzerland, 2001.

The Transportation Demand Forecasting Committee peer-reviewed this paper.

Calibrating Activity-Based Models with External Origin–Destination Information

Overview of Possibilities

Mario Cools, Elke Moons, and Geert Wets

Many practitioners question the advantages of activity-based models over conventional four-step models in regard to replication of traffic counts. This paper highlights a framework that actively links travel demand models in general and activity-based models in particular with traffic counts. Two approaches are presented that calibrate activity-based models with traffic count—an indirect and a direct approach. The indirect approach tries to incorporate findings based on the analysis of traffic counts into the components of the activity-based models. The direct approach calibrates the parameters of the travel demand model in such a way that the model replicates the observed traffic counts (quasi-) perfectly. A practical example is provided to illustrate the direct approach. The study area for this practical example is Hasselt, Belgium, a city of about 70,000 residents, and its surrounding municipalities. The practical examples revealed not a single roadway to success in calibrating activity-based models, but different options exist in fine-tuning the activity-based model. It is important to recognize some open issues and avenues for further research. First, it is not always appropriate to assume that traffic counts are completely correct. Setting up some belief structure might increase the responsiveness of the activity-based model. In addition, the origin–destination matrix calibration that optimizes the correspondence between estimated and observed screen-line counts could negatively affect the correspondence to other measures, such as vehicle miles traveled. To conclude, the formulation of a multiobjective calibration method is a key challenge for further research.

Because of an increased environmental awareness, current travel demand models pursue higher levels of behavioral realism. Four periods can be distinguished in this evolution of travel demand modeling approaches. The first period, the late 1950s, is typified by a steep increase in car use. During this period trip-based models were developed to make long-term projections of travel demand to assess major investments in road infrastructure. These first-generation models assumed that travel is the result of four consecutive steps, namely, trip generation, trip distribution, mode choice, and route choice (*1*). From the mid-1970s until the 1990s, the focus shifted toward the travel needs of a single person. The original four-step models were replaced by theories about utility maximizing behavior and individual choice behavior. Discrete choice models such as multinomial logit models and more advanced statistical techniques formed the core of so-called tour-based systems (*2*). From the mid-1990s and early 2000s activity-based travel demand models became a rising modeling paradigm. The basic premise of these third-generation models is the fact that travel behavior is a derivative of the activities that an individual performs (*1*). Current dynamic activity-based models, such as Aurora and Feathers, taking into account different forms of learning could be seen as a fourth generation of travel demand models (*3*).

Although modern activity-based travel demand models have clear theoretical advantages over conventional four-step models—the most important ones are the fact that all basic travel decisions can be applied in a disaggregate fashion, the explicit linkages that exist between travel decisions of members of a single household, the consistent choices for a single person across all travel decisions, and the disaggregate way of handling the time-of-day of travel decisions—conventional models still dominate the travel demand modeling paradigm (*4, 5*). Davidson et al. highlighted several reasons that explain the acceptance of and resistance to more sophisticated model frameworks (*6*). They can be categorized broadly as the degree of resistance to new modeling technology and the size of encouragement forces. The reasons include the size of the public agency, the size of the jurisdiction, the level of institutional history, and the level of state support for travel demand forecasting. Davidson et al. also stressed that to reinforce the transition from conventional models toward activity-based models, it is imperative that the objective theoretical advantages of activity-based models be better explained to practitioners and communicated more actively (*6*).

This paper focuses on a concern that stems from misunderstanding and mistrust by practitioners. Although researchers have acknowledged the advantages to policy analysis of an exhibited behavioral realism, many practitioners question the advantages of activity-based models over conventional four-step models in regard to replication of traffic counts because it is in many respects easier to adjust a conventional travel demand model to fit base-level traffic counts exactly than an activity-based microsimulation model (*6*). In that respect, it is important to stress the distinction between static model accuracy, in regard to the replication of the base-year observed data, and the responsive properties of the model related to the quality of the travel forecasts for future and changed conditions, because these two model properties do not necessarily coincide. Therefore, in this paper, different techniques are highlighted that actively link activity-based models in particular, and travel demand models in general, with traffic counts to achieve the desired responsive properties (the model being sensitive to demographic changes and policy measures)

Transportation Research Institute, Hasselt University, Wetenschapspark 5, Box 6, BE-3590 Diepenbeek, Belgium. Corresponding author: G. Wets, geert.wets@uhasselt.be.

Transportation Research Record: Journal of the Transportation Research Board, No. 2175, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 98–110.
DOI: 10.3141/2175-12

of the travel demand models as well as the replication of traffic counts. Proper calibration is a crucial step in simulation models because findings based on inappropriately calibrated models could be misleading and even erroneous (*7*). An overview of new calibration and validation standards, as well as best practice examples for travel demand modeling, is provided by Schiffer and Rossi (*8*). Bear in mind that the calibration of an activity-based model is not unlike calibrating a conventional four-step model (*5*). A thorough example of the calibration of a conventional four-step model with traffic counts is provided by Cascetta and Russo (*9*). For an excellent example of the calibration of an activity-based travel demand model (i.e., the Sacramento activity-based travel demand model) see Bowman et al. (*10*).

The rest of the text is organized as follows. An outline is provided of the suggested techniques, which are implemented in a practical example, followed by a thorough discussion. Finally, some general conclusions and avenues for further research are indicated.

LINKAGES BETWEEN ACTIVITY-BASED MODELS AND TRAFFIC COUNTS

There are two possible approaches to link activity-based models in particular, and travel demand models in general, with traffic counts, namely, an indirect and a direct approach. The first approach tries to incorporate findings based on the analysis of traffic counts into the model components of the activity-based models. The second approach calibrates the model parameters of the activity-based model in such a way that the model replicates the observed traffic counts (quasi-) perfectly (less than 5% error on average). The following subsections will elaborate and further clarify the two methods of linking activity-based models with traffic counts.

Indirect Linkage

The indirect linkage approach attempts to identify events that affect travel behavior and resulting traffic patterns. Analysis of traffic counts for instance can be used to identify effects of holidays and weather events (*11*). These traffic-swaying events can then be used to alter the impedance functions used in route choice modules. When events such as holidays and weather conditions are identified, their impact on travel behavior can be even further elucidated by analyzing activity diary data. Utility functions that express the propensity of performing certain activities (basically the utility functions of all elements of the activity-pattern generation can be modified in this way) can then explicitly incorporate explanatory variables to account for the events that were analyzed. In that regard, activity diary collection tools that integrate geographical information logging, such as the PARROTS tool, provide the required data to perform detailed analysis, for instance, on route choice (*12*). It can be expected that the explicit incorporation of events that account for the variability in revealed traffic patterns and their underlying reasons will result in an improved responsiveness of the activity-based model and a better replication of traffic counts.

Direct Linkage

The direct linkage approach attempts to fine-tune the model parameters of the activity-based (AB) model in such a way that the model-based traffic counts correspond maximally to the observed counts on the network. Calibration opportunities exist at four levels (Figure 1): the data level, model level, origin–destination (O-D) matrix level, and assignment level.

Two approaches can be followed when considering calibration at the data level: a "crude" approach, in which data (personal/household information, zonal information) are altered to achieve a better correspondence to the benchmark measures, and a "fine" approach, in which agents (individuals or households) are weighted. The first approach immediately raises questions concerning the validity and credibility; adjusting fields or adding or deleting records undermines the validity of the model and should be avoided. The second approach attributes weights to the different agents. For the practical example discussed in the next section, the weights are chosen to be natural numbers (including zero) such that these weights correspond to exact counterparts in the real population. Fractional weights such as 0.8 or 1.2 would also have been feasible, but the interpretation of these weights would be a probability of this agent having an exact counterpart in the real population (0.8 would correspond to a change of 80% of having an exact counterpart in the real population, and 1.2 would be interpreted as an 80% chance of having one counterpart in the real population, 20% of having two counterparts in the real population). The use of weights can be justified by the fact that there exist groups of individuals with similar travel behavior that can be captured in representative activity patterns (RAPs). By using these RAPs, the complete activity generation can be performed in a hands-on manner (*13*). McNally (*14*) and Wang (*15*) have even further advocated the use of RAPs by showing that RAPs are relatively stable over conventional planning horizons (up to 10 years). Weighting agents thus seems to be a worthwhile path to follow. Notwithstanding, the weighting procedure can become computationally very intensive as the number of possible weights increases with the number of simulated agents.

A second calibration possibility arises at the model level. The activity schedule generation could be altered in such a way that the obtained O-D matrix optimally reproduces the observed traffic counts. One solution to achieve this optimal state is an "updating" process that alters the scheduling rules derived from the available travel survey data. In addition, zone-specific rules can be introduced, for instance, increasing the probability of certain destination choices or increasing the probability of performing a certain activity. In that way, the production and attraction of these zones can be fine-tuned. When different forecasting scenarios are desired, it is necessary to keep the updated rules that were defined by the updating process in the baseline year. In that manner the AB model is constructed in a

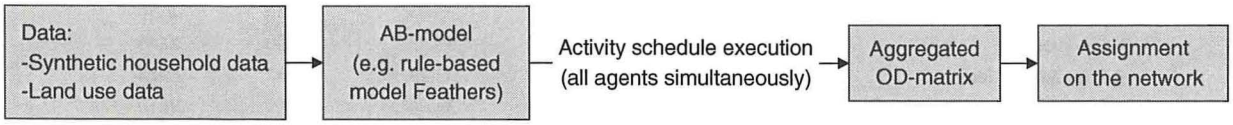


FIGURE 1 Four levels of calibration opportunities.

consistent way. Hence, linking activity-based models with traffic counts by making behavioral adjustments (altering rules) might prove to be a valid way of overcoming practitioners' mistrust.

The O-D-matrix level is the third level at which calibration opportunities arise. The O-D matrix is obtained by the simultaneous activity schedule execution of all agents. This O-D matrix can then be benchmarked (altered) in such a way that the O-D matrix better corresponds to the screen-line counts. Different techniques exist to estimate O-D matrices from traffic counts. In practice, most models require that a target O-D matrix be available or assume its availability. This target O-D matrix (the O-D matrix resulting from the AB model) is a crucial part of previous information. In statistical approaches, the target O-D matrix typically is assumed to stem from a sample survey and is regarded as an observation of the "true" O-D matrix. The observed set of traffic count data may also be assumed to be an observation of the "true" traffic count data, and therefore (small) deviations between estimated counts and observed counts may be accepted. Thus, the purpose of the calibration process is to find an O-D matrix that produces small differences between the estimated link flows and the observed flows. Three modeling philosophies are postulated in the transportation literature: traffic modeling–based approaches, statistical inference approaches, and gradient-based solution techniques (*16*).

The traffic assignment module is the last level at which calibration is possible. Obviously the way of attributing O-D flows to the network plays a crucial role in how well the model-based traffic counts correspond to the benchmark measures. Ortúzar and Willumsen classify traffic assignment methods according to their treatment of congestion (inclusion of capacity restraints) and their treatment of differences in objectives and perceptions of agents (inclusion of stochastic effects) (*17*).

PRACTICAL EXAMPLE

In this section a numerical example is provided to further illuminate the "direct linkage" approach. The study area for this numerical example is Hasselt, a Belgian city of about 70,000 residents, and its surrounding municipalities. Activity-travel information derived from census data, from the Flemish travel survey and from the O-D matrix assigned in the multimodal travel demand model, Flanders is combined to generate a simulated "true" population and its corresponding travel behavior. The data from this true population are assumed to be unbiased and precise. For generating the true RAPs at the population level, people are supposed to perform activities in a predefined order: first, people perform a work or school activity, then they go shopping, afterward they perform a leisure trip, and finally, they perform other types of activities. In addition to this predefined order, it is presumed that people perform a specific type of activity at most once (the exact chances to perform a specific activity are given in the upper part of Table 1). Furthermore, it is assumed that residents return home after their last activity.

To focus on the general ideas behind the different calibration techniques presented and to reduce model complexity route choice modeling (traffic assignment) and mode choice modeling were not taken into account. Thus, the practical example focuses on the first three levels of calibration. Assuming perfect knowledge about these aspects procures the property that the quality of the output of the (AB) travel demand model is completely related to the aggregated O-D matrix resulting from the individual activity patterns. In addition, owing to the perfect knowledge of these aspects, traffic counts on the different roads form an identity match to the O-D flows. The assumption of perfect knowledge about O-D relationships nowadays becomes a more viable option. When privacy issues are explicitly addressed, data from a mobile phone network can be

TABLE 1 Number of Residents and Propensities of Activity Participation

Municipality	No. of Residents	% Work	% School	% Shopping	% Leisure	% Other
True Population						
1. Hasselt	70,584	29.59	14.22	33.28	27.94	25.92
2. Diepenbeek	17,874	34.30	30.49	30.47	25.30	23.47
3. Kortessem	8,153	33.83	16.39	33.88	19.66	21.59
4. Alken	11,090	27.92	17.60	37.76	25.35	24.82
5. Nieuwerkerken	6,685	28.02	17.71	41.74	22.03	22.11
6. Herk-De-Stad	11,874	32.52	21.32	35.20	21.10	23.61
7. Lummen	13,874	31.38	16.38	37.03	21.19	21.77
8. Heusden-Zolder	31,017	24.54	17.95	32.19	24.18	23.63
9. Zonhoven	20,060	30.06	17.57	31.42	24.32	24.79
10. Genk	64,095	25.35	18.32	28.77	25.85	23.79
Sample Population						
1. Hasselt	1,765	28.90	13.88	32.52	30.71	25.89
2. Diepenbeek	447	35.35	27.07	30.43	23.27	23.94
3. Kortessem	204	35.78	14.22	31.37	25.98	21.57
4. Alken	277	29.96	18.41	36.82	24.55	23.83
5. Nieuwerkerken	167	23.95	20.36	40.72	23.35	20.96
6. Herk-De-Stad	297	32.32	18.86	36.03	21.55	22.56
7. Lummen	347	28.82	20.46	35.73	24.50	20.46
8. Heusden-Zolder	775	24.13	16.52	33.29	25.81	21.16
9. Zonhoven	502	30.88	22.71	34.06	26.89	26.10
10. Genk	1,602	27.47	17.04	28.34	25.22	23.78

used to derive O-D patterns (*18*). Results from Caceres et al. (*19*) and González et al. (*20*) indicate that extracting O-D information from mobile phone records has great potential and is much more cost-efficient than that generated with traditional techniques.

Because complete information about all activity patterns seldom is available, the starting point for the calibration exercises is a 2.5% stratified random sample of the "true" population (municipality is taken as the stratification variable). The lower part of Table 1 provides more information about the 2.5% sample: the number of residents in each municipality, as well as the municipality specific propensities to perform different activities, are displayed.

Table 2 presents the O-D matrix obtained from aggregating the individual activities from all people in the population (upper part of Table 2) and the sample (lower part of the table). The O-D information from the sample is scaled up to the population level for comparison purposes. A side note has to be made concerning the true population O-D matrix. When the O-D flows of this matrix are compared with flows actually observed in practice, the population O-D matrix overestimates the flows observed in practice. The reason for this overestimation is that all residents from the municipalities in this practical example are assumed to perform their activities within the entire study area.

The absolute percentage error (APE) between the true population and the sample is displayed in the lower part of Table 2. Many of these APE values are larger than 5%, indicating that some extra calibration is needed to improve the correspondence with the true observed values. The absolute percentage is defined as

$$\text{APE} = \begin{cases} \dfrac{\text{abs}\left(T_{ij}^{\text{pop}} - T_{ij}^{\text{sa}}\right)}{T_{ij}^{\text{pop}}} & \text{if } T_{ij}^{\text{pop}} > 0 \\ 0 & \text{if } T_{ij}^{\text{pop}} = T_{ij}^{\text{sa}} = 0 \\ \text{infinity value} & \text{if } T_{ij}^{\text{pop}} = 0, T_{ij}^{\text{sa}} > 0 \end{cases}$$

TABLE 2 O-D Matrices Retrieved from "True" Population and Sample

	To									
From	1	2	3	4	5	6	7	8	9	10
True Population										
1	130,888	8,142	2,692	8,239	2,620	4,580	2,899	5,270	7,108	8,928
2	8,299	22,167	1,292	744	163	283	306	825	1,281	7,682
3	2,715	1,310	8,704	522	111	106	88	137	227	1,125
4	8,278	731	518	11,780	723	656	273	322	515	673
5	2,591	151	117	721	7,052	1,683	305	219	184	335
6	4,637	318	109	648	1,614	14,892	1,852	732	316	555
7	2,891	304	95	260	308	1,907	19,398	3,272	652	783
8	5,220	837	149	314	232	721	3,281	46,967	2,953	1,960
9	7,224	1,290	241	530	175	311	673	2,915	22,160	5,621
10	8,623	7,792	1,128	711	360	534	795	1,975	5,744	112,725
Sample Population										
1	132,800	8,440	3,120	7,600	2,440	4,960	3,120	5,440	7,440	8,240
2	8,520	21,680	800	840	160	440	320	1,160	1,160	7,160
3	3,040	880	9,480	640	40	200	120	280	320	1,160
4	7,920	840	600	11,240	600	600	240	200	400	640
5	2,560	160	40	680	7,000	1,520	240	280	320	280
6	4,800	440	200	600	1,600	14,280	1,720	600	240	680
7	3,200	360	160	240	320	1,760	19,560	2,720	480	1,160
8	5,240	1,160	240	120	240	600	2,880	46,200	3,760	2,000
9	7,560	1,200	400	640	320	240	520	3,400	23,400	6,040
10	7,960	7,080	1,120	680	360	560	1,240	2,160	6,200	112,920
Absolute Percentage Difference										
1	1.5	3.7	15.9	7.8	6.9	8.3	7.6	3.2	4.7	7.7
2	2.7	2.2	38.1	12.9	1.8	55.5	4.6	40.6	9.5	6.8
3	12.0	32.8	8.9	22.6	64.0	88.7	36.4	104.4	41.0	3.1
4	4.3	14.9	15.8	4.6	17.0	8.5	12.1	37.9	22.3	4.9
5	1.2	6.0	65.8	5.7	0.7	9.7	21.3	27.9	73.9	16.4
6	3.5	38.4	83.5	7.4	0.9	4.1	7.1	18.0	24.1	22.5
7	10.7	18.4	68.4	7.7	3.9	7.7	0.8	16.9	26.4	48.2
8	0.4	38.6	61.1	61.8	3.5	16.8	12.2	1.6	27.3	2.0
9	4.7	7.0	66.0	20.8	82.9	22.8	22.7	16.6	5.6	7.5
10	7.7	9.1	0.7	4.4	0.0	4.9	56.0	9.4	7.9	0.2

where

T_{ij} = number of trips from municipality i to municipality j,
pop = indicates that flow corresponds to population, and
sa = flow corresponds to sample.

A possible infinity value could be 1, indicating being off the target by 100%. Such an infinity value has to be defined because many calculations are infeasible when values are divided by 0 (and thus mathematically are equal to infinity). Because the true population O-D matrix contains no zero cells, no infinity value had to be defined in the practical example.

Calibration at Data Level

The goal of weighting agents is to procure the highest possible resemblance between the observed traffic counts on the network and the predicted traffic counts by the activity-based model. In the noncalibrated model all agents are equally weighted (weights equal to the inverse of the sample size). By iteratively altering the weights, an optimal correspondence can be found using metaheuristics (a metaheuristic is a general algorithmic framework that can be used to guide heuristic methods to search for feasible solutions to different optimization problems). Two different approaches can be distinguished when agents have to be weighted. The first approach weights the agents before their activity pattern is generated. Because agents are duplicated before the activity patterns are generated, the activity patterns of the replicated agents—created by the weights—can differ from those of the "true" agents. Thus, the convergence of the iterative process of weighting persons and calculating the activity patterns of the "agents" and their replicates is not necessarily guaranteed. The second approach solves this convergence problem by weighting the activity patterns instead of the agents themselves. Take for example a resident in Hasselt who performs a work activity only in Diepenbeek. From Table 2 it can be seen that if this person's weight would be decreased, the estimated O-D flows from Hasselt to Diepenbeek and Diepenbeek to Hasselt would be reduced, and thus would be closer to the "true" O-D flows for the population.

To illustrate the calibration of O-D matrices at the data levels, the second approach, the weighting of activity patterns, is followed. The RAPs of the residents in the sample are weighted by using the algorithm displayed in Figure 2. The algorithm that is implemented includes an element originating from tabu search metaheuristics, namely, the concept of a tabu list. A tabu list is a short-term memory in which, in this case, the persons whose weights have been altered are stored (*21*). The tabu list ensures that these weights are not altered multiple times in the same iteration, thus preventing situations such as the repetitive increasing and decreasing of the weight of a specific person. Two versions of the algorithm were implemented. The first one changed the weights by adding or subtracting one. The second one altered the weights by increasing or reducing the weights by a random number between one and 10, reducing the risk of converging toward the same saddle point [i.e., the same (sub)optimum]. A safeguard was included, procuring nonnegative weights.

The estimated O-D matrices are provided in Table 3. The mean APE (MAPE) of the estimated matrix using the first algorithm equals 2.12%; the second matrix has a MAPE of 2.02%. From Table 4 it can be seen that for two cells in both matrices the APE is higher than 0.5. The reason is that very few people are traveling between these two locations (Kortessem and Nieuwerkerken), and in line with this, that the people in the sample traveling between these locations, also travel between other uncommon O-D pairs (Kortessem–Herk-De-Stad and Kortessem–Lummen). This underlines the importance of including a stop criterion in the algorithms to avoid an endless computation.

Calibration at Model Level

The basic model that will be calibrated first predicts activity chains for all persons (the proportions of the different activity chains have been fixed to the population proportions) and then predicts the locations where the different activities will be performed. The proportions of the different activity chains have been fixed to the population proportions. This ensures that discrepancies between the "true population" O-D matrix and the calibrated O-D matrix are due only to differences in destination choices (location probabilities). Thus, at the model level, the activity schedule generation could be altered by iteratively updating the probabilities of certain destination choices (related to their respective activity purposes). The adjustment of the model parameters is straightforward in this case because only one dimension is considered at a time (i.e., the location probabilities). After all, the other parameters (such as the chances of performing certain activities) are kept constant. For real AB models in practice, a chain of interlinked choices with feedbacks are modeled, and thus multiple parameters have to be changed simultaneously. This would seriously augment the complexity of the model, but the basic framework elucidated in this paper still could be used.

The updating process will attain a quasi-perfect match when the updated sample probabilities of the destination choices are equal to the unknown population probabilities. Nonetheless, a full search of the solution space (investigating all possible combinations of location probabilities for the different activities) is not a feasible option because the number of possible combinations approaches infinity. The number of possible combinations can be computed as follows:

$$\left(\frac{1}{(1-\text{precision of location probability})}\right)^{\text{number of activities} \times \text{number of municipalities}^2}$$

which, for the practical example discussed in this paper (applying a precision of 1%), would yield a total number of possible combinations of 10^{500} (approximating infinity). Therefore, an algorithm that explores the solution space for a "good" solution instead of the optimal solution should be implemented.

To calibrate the AB travel demand model and to ensure convergence of optimization algorithms, it is essential that the variability caused by the activity-generation process be reduced as much as possible. Stability of the activity generation can be ensured by taking averages over multiple (activity generation) runs, so that differences between the estimated O-D matrix and the true population O-D matrix are not the result of random variations, but of the altered location probabilities. However, guaranteeing the stability of the activity generation diminishes the performance because computation times are significantly increased. The algorithm used is shown in Figure 3.

Table 5 presents the O-D matrix and corresponding APEs for the model-based calibration results. From these results, one can see that here is a decrease in the MAPE from 20.27% in the upscaled sample O-D matrix to 6.29% in the model-calibrated O-D matrix (after 100 iterations). Nevertheless, because multiple activity generations

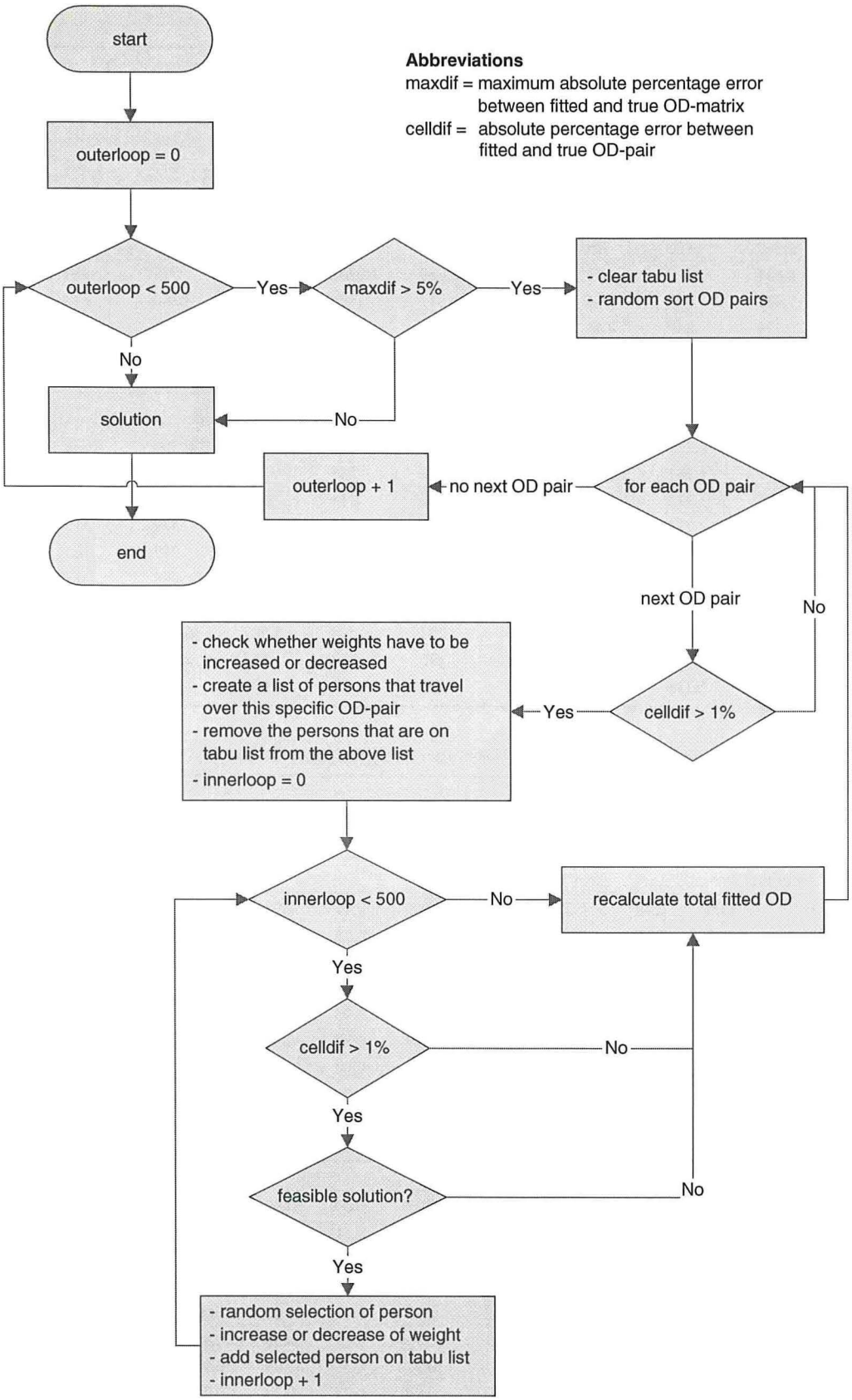


FIGURE 2 Calibration algorithm to weight representative activity patterns.

TABLE 3 O-D Matrices Calibrated Using Weighted RAPs

	To									
From	1	2	3	4	5	6	7	8	9	10
Algorithm 1										
1	132,196	8,181	2,702	8,194	2,594	4,608	2,871	5,298	7,044	8,855
2	8,291	21,972	1,282	738	160	285	304	817	1,269	7,751
3	2,688	1,319	8,781	526	32	107	88	138	225	1,130
4	8,241	738	513	11,715	716	650	271	325	517	670
5	2,567	160	46	727	7,038	1,667	302	217	185	338
6	4,611	315	107	654	1,630	14,762	1,842	739	314	559
7	2,915	307	95	262	311	1,896	19,585	3,240	648	777
8	5,179	845	148	311	234	714	3,309	46,563	2,959	1,955
9	7,263	1,302	242	525	174	314	676	2,887	22,286	5,565
10	8,592	7,730	1,118	704	358	530	788	1,993	5,787	113,222
Algorithm 2										
1	132,173	8,223	2,712	8,160	2,610	4,584	2,874	5,306	7,038	8,873
2	8,332	21,987	1,282	744	160	285	304	818	1,284	7,720
3	2,695	1,323	8,728	526	40	111	88	138	226	1,120
4	8,227	724	514	11,814	718	650	271	324	519	667
5	2,601	160	40	718	6,985	1,670	303	219	185	337
6	4,610	315	111	654	1,630	14,906	1,859	729	313	557
7	2,905	304	95	262	311	1,920	19,570	3,240	657	779
8	5,182	844	148	314	232	714	3,307	46,870	2,966	1,942
9	7,273	1,302	239	527	175	313	667	2,896	22,292	5,565
10	8,555	7,734	1,126	709	357	531	800	1,979	5,769	113,457

TABLE 4 Absolute Percentage Errors (Calibrated Using Weighted RAPs)

	To									
From	1	2	3	4	5	6	7	8	9	10
Algorithm 1										
1	1.00	0.48	0.37	0.55	0.99	0.61	0.97	0.53	0.90	0.82
2	0.10	0.88	0.77	0.81	1.84	0.71	0.65	0.97	0.94	0.90
3	0.99	0.69	0.88	0.77	71.17	0.94	0.00	0.73	0.88	0.44
4	0.45	0.96	0.97	0.55	0.97	0.91	0.73	0.93	0.39	0.45
5	0.93	5.96	60.68	0.83	0.20	0.95	0.98	0.91	0.54	0.90
6	0.56	0.94	1.83	0.93	0.99	0.87	0.54	0.96	0.63	0.72
7	0.83	0.99	0.00	0.77	0.97	0.58	0.96	0.98	0.61	0.77
8	0.79	0.96	0.67	0.96	0.86	0.97	0.85	0.86	0.20	0.26
9	0.54	0.93	0.41	0.94	0.57	0.96	0.45	0.96	0.57	1.00
10	0.36	0.80	0.89	0.98	0.56	0.75	0.88	0.91	0.75	0.44
Algorithm 2										
1	0.98	0.99	0.74	0.96	0.38	0.09	0.86	0.68	0.98	0.62
2	0.40	0.81	0.77	0.00	1.84	0.71	0.65	0.85	0.23	0.49
3	0.74	0.99	0.28	0.77	63.96	4.72	0.00	0.73	0.44	0.44
4	0.62	0.96	0.77	0.29	0.69	0.91	0.73	0.62	0.78	0.89
5	0.39	5.96	65.81	0.42	0.95	0.77	0.66	0.00	0.54	0.60
6	0.58	0.94	1.83	0.93	0.99	0.09	0.38	0.41	0.95	0.36
7	0.48	0.00	0.00	0.77	0.97	0.68	0.89	0.98	0.77	0.51
8	0.73	0.84	0.67	0.00	0.00	0.97	0.79	0.21	0.44	0.92
9	0.68	0.93	0.83	0.57	0.00	0.64	0.89	0.65	0.60	1.00
10	0.79	0.74	0.18	0.28	0.83	0.56	0.63	0.20	0.44	0.65

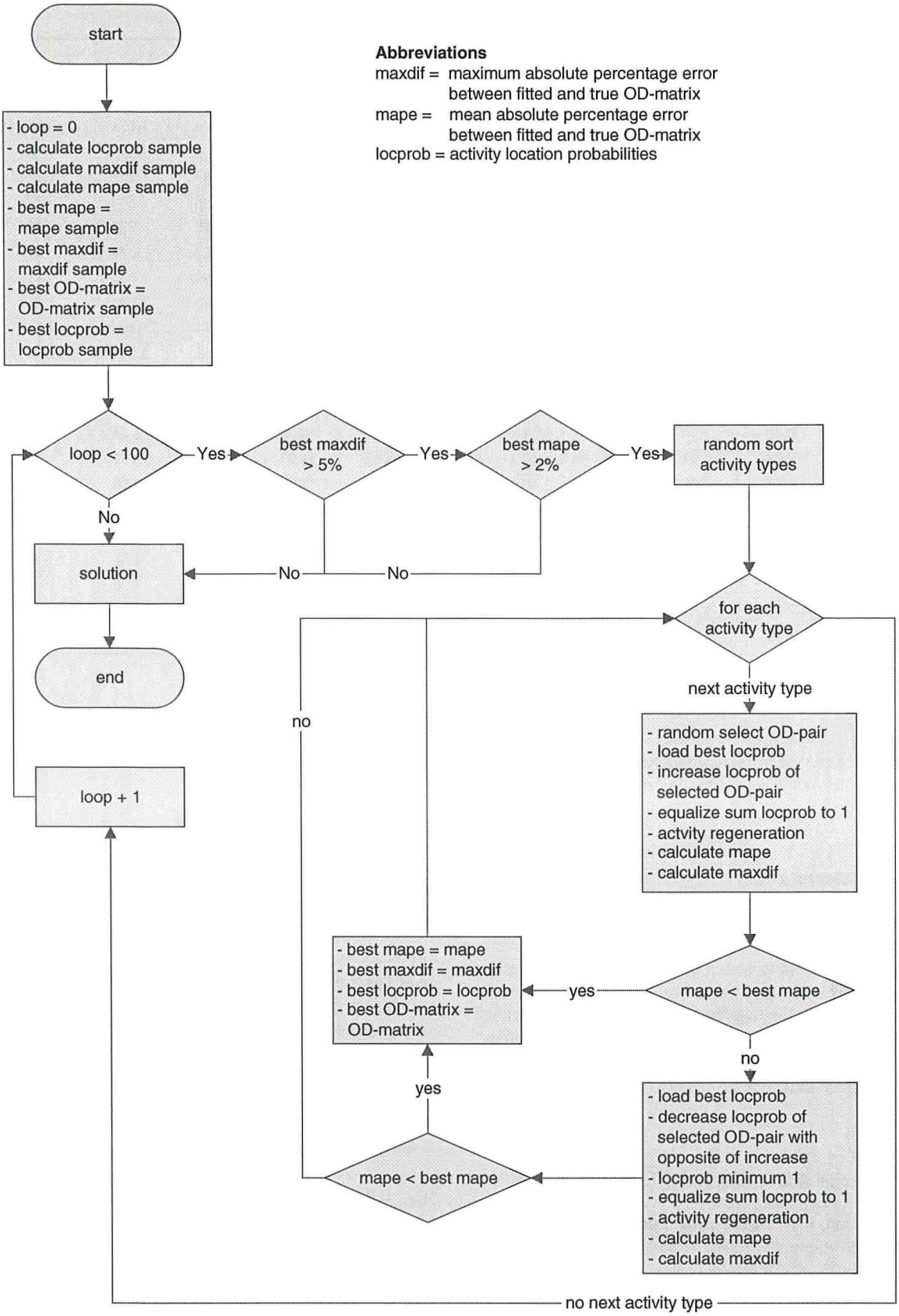


FIGURE 3 Calibration algorithm to adjust activity location probabilities.

TABLE 5 Model-Based Calibrated O-D Matrix and Corresponding APEs

	To									
From	1	2	3	4	5	6	7	8	9	10
Model-Based Calibrated O-D Matrix										
1	131,667	8,088	2,822	8,140	2,553	4,634	2,970	5,391	7,071	8,617
2	8,307	21,901	1,153	783	178	319	331	909	1,211	7,554
3	2,830	1,188	8,814	554	98	126	107	183	258	1,174
4	8,159	759	543	11,487	690	623	234	265	544	711
5	2,506	172	101	689	7,063	1,665	303	236	220	355
6	4,692	341	124	628	1,591	14,879	1,815	682	324	615
7	2,972	333	107	228	312	1,843	19,266	3,149	635	857
8	5,306	918	190	263	235	685	3,188	46,990	3,064	2,044
9	7,154	1,241	269	545	224	321	631	3,014	21,894	5,772
10	8,359	7,706	1,208	699	367	597	856	2,063	5,844	112,689
Model-Based Calibrated APEs										
1	0.60	0.66	4.84	1.21	2.54	1.17	2.44	2.30	0.52	3.48
2	0.10	1.20	10.78	5.24	9.20	12.84	8.28	10.18	5.44	1.66
3	4.24	9.34	1.27	6.19	11.41	18.55	21.97	33.58	13.51	4.33
4	1.43	3.88	4.83	2.49	4.56	5.08	14.29	17.70	5.63	5.65
5	3.27	14.13	13.68	4.48	0.16	1.05	0.77	7.76	19.38	5.97
6	1.19	7.34	14.07	3.09	1.45	0.09	2.00	6.83	2.64	10.81
7	2.79	9.43	12.28	12.44	1.19	3.36	0.68	3.75	2.66	9.45
8	1.65	9.64	27.74	16.14	1.15	5.04	2.84	0.05	3.76	4.27
9	0.96	3.80	11.62	2.77	27.81	3.22	6.29	3.38	1.20	2.69
10	3.06	1.11	7.12	1.69	1.85	11.86	7.67	4.46	1.74	0.03

NOTE: APE = absolute percentage error.

are required in each step of the algorithm, model-based calibration is the most computer-intensive calibration option, favoring other calibration techniques.

Calibration at Matrix Level

The third level of calibration tackled in this study is the matrix level. Recall that perfect knowledge about route choice and mode choice is assumed and that an identity match is presumed between traffic counts and O-D flows. Therefore the calibration at the matrix level, as in the two previous calibration levels discussed, is illustrated by using O-D-pair information. [See Abrahamsson for a thorough literature review concerning the calibration of O-D matrices using traffic counts (*16*).] Three situations are explored to calibrate the survey O-D matrix.

Perfect Knowledge About Interzonal Traffic

In the first situation, it is assumed that "perfect" knowledge is available about all interzonal traffic flows, but that information about intrazonal traffic is available only at the survey level. Let $P_i = \Sigma_j T_{ij}$ be the number of trips originating from municipality i (production), $A_j = \Sigma_i T_{ij}$ be the number of trips arriving in municipality j (attraction), and T_{ij} be the number of trips from zone i to zone j.

Then the intrazonal traffic flows ($T_{ij,i=j}$) could be approached by the following formula:

$$T_{ij,i=j} = \lambda\left(P_i^{\text{est}} - P_i^{*,\text{pop}}\right) + (1-\lambda)\left(A_j^{\text{est}} - A_j^{*,\text{pop}}\right)$$

where $\lambda \in [0, 1]$ expresses the relative importance that is given to the number of trips originating in a municipality, compared with the number of trips arriving in a municipality, where est indicates that the quantity is derived from the estimated (survey) O-D matrix, and pop indicates that the quantity is derived from the population "true" O-D matrix. The asterisk underlines the fact that the intrazonal traffic flows are not included in the population row ($P_i^{*,\text{pop}} = \Sigma_{j,j\neq i} T_{ij}^{\text{pop}}$) and column totals ($A_j^{*,\text{pop}} = \Sigma_{i,i\neq j} T_{ij}^{\text{pop}}$). As it is often assumed that production is estimated more accurately than attraction (*17*), in this practical example three times more confidence is placed in the estimation of productions than in the estimation of attractions. Thus, the intrazonal O-D flows are calculated as follows:

$$T_{ij,i=j} = 0.75\left(P_i^{\text{est}} - P_i^{*,\text{pop}}\right) + 0.25\left(A_j^{\text{est}} - A_j^{*,\text{pop}}\right)$$

The resulting O-D matrix is given in the upper part of Table 6. When it is assumed that the activity-travel pattern of people begins and ends in the home location (as is the case for the practical applications described in this paper), the number of trips originating from a municipality equals the number of trips arriving in that municipal-

TABLE 6 O-D Matrices Calibrated by Using Matrix-Level Possibilities

	To									
From	1	2	3	4	5	6	7	8	9	10
Situation 1. Perfect Knowledge About Interzonal Traffic										
1	133,122	8,142	2,692	8,239	2,620	4,580	2,899	5,270	7,108	8,928
2	8,299	21,365	1,292	744	163	283	306	825	1,281	7,682
3	2,715	1,310	9,819	522	111	106	88	137	227	1,125
4	8,278	731	518	10,591	723	656	273	322	515	673
5	2,591	151	117	721	6,774	1,683	305	219	184	335
6	4,637	318	109	648	1,614	14,379	1,852	732	316	555
7	2,891	304	95	260	308	1,907	19,488	3,272	652	783
8	5,220	837	149	314	232	721	3,281	46,773	2,953	1,960
9	7,224	1,290	241	530	175	311	673	2,915	24,740	5,621
10	8,623	7,792	1,128	711	360	534	795	1,975	5,744	112,618
Situation 2. Growth Factor Modeling (Furness iteration)										
1	132,854	8,085	2,839	8,008	2,606	4,556	2,925	5,294	7,449	8,984
2	8,241	21,536	1,333	707	159	275	302	811	1,313	7,563
3	2,863	1,352	9,535	527	115	110	92	143	247	1,176
4	8,046	695	523	10,964	688	625	264	310	517	648
5	2,577	147	121	687	6,871	1,640	302	216	189	330
6	4,611	309	113	617	1,573	14,513	1,831	720	325	548
7	2,919	301	100	251	305	1,886	19,468	3,268	679	783
8	5,243	822	155	302	228	710	3,278	46,688	3,062	1,952
9	7,569	1,322	262	532	180	319	702	3,023	23,972	5,839
10	8,677	7,671	1,179	685	355	526	796	1,967	5,967	112,457
Situation 3. Perceived Precision Updating										
1	131,207	8,192	2,763	8,132	2,590	4,643	2,936	5,298	7,163	8,813
2	8,336	22,086	1,210	760	162	309	308	881	1,261	7,595
3	2,769	1,238	8,833	542	99	122	93	161	242	1,131
4	8,218	749	532	11,690	702	647	267	302	496	667
5	2,586	152	104	714	7,043	1,656	294	229	207	326
6	4,664	338	124	640	1,612	14,790	1,830	710	303	576
7	2,942	313	106	257	310	1,882	19,425	3,180	623	846
8	5,223	891	164	282	233	701	3,214	46,839	3,087	1,967
9	7,280	1,275	267	548	199	299	647	2,996	22,367	5,691
10	8,512	7,673	1,127	706	360	538	869	2,006	5,820	112,758

ity. In this case the choice of λ is irrelevant. From Table 7 it is clear that only the intrazonal trips are altered (APEs for interzonal trips equal zero).

Growth Factor Modeling (Furness Iteration)

The second situation considers the case in which two O-D matrices (one on the population level and one derived from the sample) are available. Information from these O-D matrices can be combined by using growth factor modeling. One option is to take the cell information from the population (e.g., retrieved from Global Positioning System tracks) and the trip totals (column and row totals of the O-D matrix) from the survey. A second option is the reverse, namely, taking the cell information from the survey and the trip totals from the population. To illustrate the technique, the first option is implemented. This option is the more realistic one, because in practice precise O-D-pair information can be derived by using cell phone information at fairly low cost, and surveys capture well the total travel demand. The doubly constrained growth factor model is estimated by using Furness iterations. Formally, the number of trips from municipality i to j (T_{ij}) is calculated as follows:

$$T_{ij} = t_{ij} \times a_i \times b_j$$

where t_{ij} is the number of trips (in the population O-D matrix) and a_i and b_j are balancing factors. These balancing factors are a set of correction coefficients that are appropriately applied to the cell entries in each row or column. The iterative procedure starts with setting all

TABLE 7 APEs (Calibrated Using Matrix Level Possibilities)

	To									
From	1	2	3	4	5	6	7	8	9	10
Situation 1. Perfect Knowledge About Interzonal Traffic										
1	1.71	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	3.62	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	12.81	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	10.09	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	3.94	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	3.44	0.00	0.00	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00	0.46	0.00	0.00	0.00
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.41	0.00	0.00
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	11.64	0.00
10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.09
Situation 2. Growth Factor Modeling (Furness iteration)										
1	1.50	0.70	5.46	2.80	0.53	0.52	0.90	0.46	4.80	0.63
2	0.70	2.85	3.17	4.97	2.45	2.83	1.31	1.70	2.50	1.55
3	5.45	3.21	9.55	0.96	3.60	3.77	4.55	4.38	8.81	4.53
4	2.80	4.92	0.97	6.93	4.84	4.73	3.30	3.73	0.39	3.71
5	0.54	2.65	3.42	4.72	2.57	2.55	0.98	1.37	2.72	1.49
6	0.56	2.83	3.67	4.78	2.54	2.54	1.13	1.64	2.85	1.26
7	0.97	0.99	5.26	3.46	0.97	1.10	0.36	0.12	4.14	0.00
8	0.44	1.79	4.03	3.82	1.72	1.53	0.09	0.59	3.69	0.41
9	4.78	2.48	8.71	0.38	2.86	2.57	4.31	3.70	8.18	3.88
10	0.63	1.55	4.52	3.66	1.39	1.50	0.13	0.41	3.88	0.24
Situation 3. Perceived Precision Updating										
1	0.24	0.61	2.65	1.29	1.15	1.38	1.27	0.54	0.78	1.28
2	0.44	0.37	6.35	2.15	0.31	9.25	0.76	6.77	1.57	1.13
3	2.00	5.47	1.49	3.77	10.66	14.78	6.06	17.40	6.83	0.52
4	0.72	2.49	2.64	0.76	2.84	1.42	2.01	6.31	3.72	0.82
5	0.20	0.99	10.97	0.95	0.12	1.61	3.55	4.64	12.32	2.74
6	0.59	6.39	13.91	1.23	0.14	0.68	1.19	3.01	4.01	3.75
7	1.78	3.07	11.40	1.28	0.65	1.28	0.14	2.81	4.40	8.02
8	0.06	6.43	10.18	10.30	0.57	2.80	2.04	0.27	4.55	0.34
9	0.78	1.16	11.00	3.46	13.81	3.80	3.79	2.77	0.93	1.24
10	1.28	1.52	0.12	0.73	0.00	0.81	9.33	1.56	1.32	0.03

b_j equal to one. In the second step, the a_i are solved for b_j to satisfy the trip production constraint (row totals of the cell entries of the population O-D matrix have to equal the productions derived from the survey). Subsequently, in the third step, the b_j are solved for the a_i, calculated in the previous step, to satisfy the trip attraction constraint (column totals of the cell entries of the population O-D matrix have to equal the attractions derived from the survey). Then, the O-D matrix is updated. This consecutive calculation of a_i and b_j is repeated until convergence is achieved (the production and attraction constraints are satisfied). The procedure yields the matrix presented in the middle of Table 6, the corresponding APEs in Table 7.

Perceived Precision Updating

The third and final situation explored to illustrate potential calibration options at the data level describes the case in which an outdated population-based O-D matrix, as well as a recent matrix derived from the sample, is available. The procedure is an adaptation of the Bayesian updating procedure discussed by Atherton and Ben-Akiva (*22*). This procedure updates information by using the following formulas:

$$\theta_{updated} = \frac{\frac{\theta_{prior}}{\sigma^2_{prior}} + \frac{\theta_{updating}}{\sigma^2_{updating}}}{\frac{1}{\sigma^2_{prior}} + \frac{1}{\sigma^2_{updating}}}$$

and

$$\sigma^2_{updated} = \frac{1}{\frac{1}{\sigma^2_{prior}} + \frac{1}{\sigma^2_{updating}}}$$

where θ is the mean of the investigated quantity and σ^2 the variance of the mean of that quantity. Because the O-D cells in an O-D matrix are fixed numbers, of which the variance is seldom reported, one could replace the mean of the quantity by the O-D flow and reformulate the formulas in relation to perceived precision (ψ) instead of variance of the mean (because the precision increases as the variance decreases). This perceived precision can for instance be obtained via expert knowledge. The formulas then take the form of the following equations:

$$T_{ij}^{new} = \frac{\frac{T_{ij}^{pop}}{\left(1-\psi^{pop}\right)} + \frac{T_{ij}^{sa}}{\left(1-\psi^{sa}\right)}}{\frac{1}{\left(1-\psi^{pop}\right)} + \frac{1}{\left(1-\psi^{sa}\right)}}$$

and

$$\psi^{new} = 1 - \frac{1}{\frac{1}{\left(1-\psi^{pop}\right)} + \frac{1}{\left(1-\psi^{sa}\right)}}$$

For the practical example discussed in this paper the perceived precision of the population O-D matrix is set equal to 99% and that of the sample O-D matrix equal to 95%. The updated O-D matrix then has a precision of 99.17%. The updated O-D matrix is shown in the lower part of Table 6. For reasons of completeness and comparability with other calibration techniques, the APEs for this method are also presented (Table 7), even though interpretation of these specific APEs is meaningless, as the premise of this example is outdated population data.

Discussion of Proposed Techniques

An interesting issue of calibration to traffic counts is the fact that traffic counts themselves are uncertain. Uncertainty can be tackled in the data-level- and model-level-based calibration by adjusting the converge criterion, that is, absolute percentage errors [denoted as fitness values by Park and Qi (*7*)]. When a choice is being made between the different techniques suggested in this paper, three key issues have to be taken into account: computational complexity, data availability, and sensitivity to policy issues.

The most computer-intensive method was the model-based calibration, requiring 14 days of computation on a computer with a Core 2 Duo 2.10 GHz CPU and 4 GB RAM. This large computation time resulted because the calibration at this level involves running the full simulation model (*23, 24*). In comparison, the iterative procedure for calibration at the data level took about 1 day, and the matrix-level techniques required only a few seconds of computation (the latter techniques did not include iterative optimization techniques). The computation times of the iterative procedures could be decreased by using more-efficient optimization algorithms, such as genetic algorithms (*7*) and golden section search (*25*).

Next to the computational complexity, the available target data will definitely influence the suitability of the different techniques. The largest amount of target data is required for the model-based calibration because for each subpart of the model, target information is necessary.

Finally, the influence of the calibration techniques on the sensitivity of the model to policy measures is of high importance. This sensitivity depends on how the base-year calibration manipulations (i.e., calibrations weights) are transferred toward future predictions. Further research on the policy sensitivity of the different approaches should be a key priority for future research.

CONCLUSIONS AND FURTHER RESEARCH

In this paper different possibilities for linking travel demand models in general, and AB models in particular, with traffic counts and precise O-D matrix information are highlighted and illustrated by means of an example. The discussed techniques provide the framework to overcome one of the main concerns of practitioners, namely, the disadvantage of AB models over conventional four-step models in regard to the replication of traffic counts. The practical examples revealed that there is not a single roadway to success in calibrating AB models, but that different options exist in fine-tuning the AB model. Therefore, a careful assessment of the available options is needed to determine which choices have to be made. A step-by-step procedure, combining elements of the different proposed solutions, can be recommended.

Notwithstanding, it is important to recognize some open issues and avenues for further research. First, it is not always appropriate to assume that traffic counts are completely correct. In reality, differences may relate to sampling bias, variability in travel, imperfect counts, assumptions about nonpassenger cars (e.g., freight traffic) and external traffic, and unreliability in model facets. Setting up some belief structure might increase the responsiveness of the AB model. Second, the O-D matrix calibration that optimizes the correspondence between estimated and observed screen-line counts could negatively affect the correspondence to other measures such as vehicle miles traveled. Thus, formulation of a multiobjective calibration method is a key challenge. Third, in most cases in practice, travel demand models are validated and tested against hour-specific counts. The same methodology can be applied in this case: modeled trip tables must be compared with counts for each time-of-day period. The challenge here exists in consolidating the time-of-day-specific adjustments into a set of activity-generation, location, and schedule adjustments. Finally, additional testing of calibration possibilities in a real AB travel demand modeling environment would provide further empirical evidence of the proposed frameworks. In particular, the investigation of how the policy sensitivity of an AB model is affected by the different approaches should be a key priority for further research.

ACKNOWLEDGMENT

The authors thank Benoît Depaire for his advice on the implementation of the calibration algorithms.

REFERENCES

1. Jovicic, G. *Activity Based Travel Demand Modelling: A Literature Study.* Note 8. Danmarks TransportForskning, Lyngby, Denmark, 2001.
2. Ben-Akiva, M., and S. R. Lerman. *Discrete Choice Analysis: Theory and Application to Travel Demand.* MIT Press, Cambridge, Mass., 1985.
3. Bellemans, T., D. Janssens, G. Wets, T. Arentze, and H. Timmermans. Implementation Framework and Development Trajectory of FEATHERS Activity-Based Simulation Platform. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2175,* Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 111–119.

4. Vovsha, P., M. Bradley, and J. Bowman. Activity-Based Travel Forecasting Models in the United States: Progress Since 1995 and Prospects for the Future. In *Progress in Activity-Based Analysis* (H. Timmermans, ed.), Elsevier Science, Ltd., Oxford, United Kingdom, 2005, pp. 389–414.
5. Walker, J. L. Making Household Microsimulation of Travel and Activities Accessible to Planners. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1931,* Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 38–48.
6. Davidson, W., R. Donnelly, P. Vosha, J. Freedman, S. Ruegg, J. Hicks, J. Castiglione, and R. Picado. Synthesis of First Practices and Operational Research Approaches in Activity-Based Travel Demand Modeling. *Transportation Research Part A,* Vol. 41, No. 5, 2007, pp. 464–488.
7. Park, B., and H. Qi. Development and Evaluation of a Procedure for the Calibration of Simulation Models. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1934,* Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 208–217.
8. Schiffer, R. G., and T. F. Rossi. New Calibration and Validation Standards for Travel Demand Modeling. Presented at 88th Annual Meeting of the Transportation Research Board, Washington, D.C., 2009.
9. Cascetta, E., and F. Russo. Calibrating Aggregate Travel Demand Models with Traffic Counts: Estimators and Statistical Performance. *Transportation,* Vol. 24, No. 3, 1997, pp. 271–293.
10. Bowman, J. L., M. A. Bradley, and J. Gibb. The Sacramento Activity-Based Travel Demand Model: Estimation and Validation Results. *Proc., European Transport Conference 2006.* Association for European Transport, Strasbourg, France, 2006.
11. Cools, M., E. A. Moons, and G. Wets. Investigating Effect of Holidays on Daily Traffic Counts: Time Series Approach. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2019,* Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 22–31.
12. Bellemans, T., B. Kochan, D. Janssens, G. Wets, and H. J. P. Timmermans. Field Evaluation of Personal Digital Assistant Enabled by Global Positioning System: Impact on Quality of Activity and Diary Data. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2049,* Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 136–143.
13. McNally, G. M., and A. A. Kulkarni. Microsimulation of Daily Activity Patterns. Presented at 80th Annual Meeting of the Transportation Research Board, Washington, D.C., 2001.
14. McNally, M. Activity-Based Forecasting Models Integrating GIS. *Geographical Systems,* Vol. 5, 1999, pp. 163–187.
15. Wang, R. *An Activity-Based Microsimulation Model.* PhD dissertation. University of California, Irvine, 1996.
16. Abrahamsson, T. *Estimation of Origin–Destination Matrices Using Traffic Counts: A Literature Survey.* IIASA Interim Report IR-98-021. May 1998.
17. Ortúzar, J., and L. Willumsen. *Modeling Transport,* 3rd ed. John Wiley & Sons, Ltd., Chichester, United Kingdom, 2001.
18. Giannotti, F., and D. Pedreschi. *Mobility, Data Mining and Privacy: Geographic Knowledge Discovery.* Springer, Berlin, 2008.
19. Caceres, N., J. P. Wideberg, and F. G. Benitez. Deriving Origin–Destination Data from a Mobile Phone Network. *IET Intelligent Transportation Systems,* Vol. 1, No. 1, 2007, pp. 15–26.
20. González, M., C. A. Hidalgo, and A.-L. Barabási. Understanding Individual Human Mobility Patterns. *Nature,* Vol. 453, 2008, pp. 779–782.
21. Glover, F. Tabu Search–Part II. *ORSA Journal of Computing,* Vol. 2, No. 1, 1990, pp. 4–32.
22. Atherton, T. J., and M. E. Ben-Akiva. Transferability and Updating of Disaggregate Travel Demand Models. In *Transportation Research Record 610,* TRB, National Research Council, Washington, D.C., 1976, pp. 12–18.
23. Jha, M., G. Gopalan, A. Garms, B. P. Mahanti, T. Toledo, and M. E. Ben-Akiva. Development and Calibration of a Large-Scale Microscopic Traffic Simulation Model. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1876,* Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 121–131.
24. Toledo, T., M. E. Ben-Akiva, D. Darda, M. Jha, and H. N. Koutsopoulos. Calibration of Microscopic Traffic Simulation Models with Aggregate Data. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1876,* Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 10–19.
25. Zhang, L., and D. Levinson. Agent-Based Approach to Travel Demand Modeling: Exploratory Analysis. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1898,* Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 28–36.

The Transportation Demand Forecasting Committee peer-reviewed this paper.

Implementation Framework and Development Trajectory of FEATHERS Activity-Based Simulation Platform

Tom Bellemans, Bruno Kochan, Davy Janssens, Geert Wets, Theo Arentze, and Harry Timmermans

To facilitate the development of dynamic activity-based models for transport demand, the FEATHERS framework was developed. This framework suggests a four-stage development trajectory for a smooth transition from the four-step models toward static activity-based models in the short term and dynamic activity-based models in the long term. The development stages discussed in this paper range from an initial static activity-based model without traffic assignment to a dynamic activity-based model that incorporates rescheduling, learning effects, and traffic routing. To illustrate the FEATHERS framework, work that has been done on the development of static and dynamic activity-based models for Flanders (Belgium) and the Netherlands is discussed. First, the data collection is presented. Next, the four-stage activity-based model development trajectory is discussed in detail. The paper concludes with the presentation of the modular FEATHERS framework, which discusses the functionalities of the modules and how they accommodate the requirements imposed on the framework by each of the four stages.

During the past decade, several microsimulation models of activity–travel demand [e.g., Cemdap (*1*), Famos (*2*), and Albatross (*3, 4*)] have become operational. That led to an increased concern to move the currently operational, and newly developed activity-based models, into practice. Especially in Europe, advanced tour-based models have already introduced some of these interdependencies and, hence, operational applications of models that involve microsimulation of activity–travel patterns have remained limited. Although several practical reasons for this slow dissemination can be envisaged, one of the main challenges faced by the travel demand forecasting industry is the ability to rapidly deploy several new theoretical advances in a time- and cost-efficient manner (*5*). Although small laboratory experiments are needed for exploring these theoretical advances from a scientific point of view, it is of utmost importance to rely on a sound basic platform in which several of these advancements can serve as add-ons if one is concerned about the final operationalization of the developed tools. A nice example of such a platform is the open source Multiagent Transport Simulation Toolkit (MATSIM-T) project, in which some basic functionality of a multiagent microsimulation for transport planning has been implemented featuring implementations of dynamic traffic network assignments (*6*).

Taking the above into account, the idea was conceived in Flanders, Belgium, to develop a modular activity-based model of transport demand, in which the emphasis is, on the one hand, on methodologically innovative (dynamic) activity–travel demand generation and, on the other hand, on the practical use of the system by practitioners and end users. The modularity of the software is ensured by design in using the object-oriented paradigm, allowing for a more flexible application programming structure.

A four-stage development trajectory has been postulated in the context of FEATHERS: Stage 1 is the development of a static activity-based model; Stage 2 is the development of a semistatic model accounting for evolutionary and nonstationary behavior (for instance, different time periods during the day, different days); Stage 3 is the development of a fully dynamic activity-based model accounting for short-term adaptation behavior and learning; and Stage 4 is the development of a full dynamic agent-based microsimulation framework involving traffic and route assignment on a microscopic level. This development trajectory is innovative because most microsimulation models of activity–travel demand are situated either in Stage 1 or in Stage 2. Indeed, in regard to short-term dynamics in activity–travel patterns and travel execution (Stage 3 and 4), activity-based models have little to offer at their current state of development. Apart from the MATSIM-T framework, the aggregate impact of individual-level route choice decisions on activity generation and rescheduling behavior is not included in activity-based models. Issues such as uncertainty, learning, and nonstationary environments are also not considered. Of course, there is a wide variety of literature available on traffic assignment, route, and departure choice models, but at their current state of development it is fair to say that the behavioral contents of these models from an activity-based perspective are still relatively weak and that comprehensive dynamic models are still lacking.

The multistage process outlined above is crucial in understanding and accounting for end user (Flemish government, environmental agencies, public transport providers) concerns, in which currently a traditional four-step modeling approach is used in Flanders. Moving directly toward a full agent-based microsimulation framework is therefore not appealing from an end user point of view given the challenges of data collection and computational complexity. Hence, the four-stage approach presented in this paper allows for a gradual evolution toward more sophisticated models as time and

T. Bellemans, B. Kochan, D. Janssens, and G. Wets, Transportation Research Institute, Faculty of Applied Economics, Hasselt University, Wetenschapspark 5, Box 6, B-3590 Diepenbeek, Belgium. T. Arentze and H. Timmermans, Urban Planning Group, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, Netherlands. Corresponding author: G. Wets, geert.wets@uhasselt.be.

Transportation Research Record: Journal of the Transportation Research Board, No. 2175, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 111–119.
DOI: 10.3141/2175-13

budget constraints permit, while aiming at maximally reusing previous efforts and investments. The research described in this paper has been given the acronym FEATHERS, and the application area is in line with other existing activity-based models but is extended toward environmental, health, and in the medium term, traffic safety applications.

To set up an activity-based microsimulation, one needs considerable amounts of data. The rest of this paper therefore first discusses an extensive hybrid multimethod data collection approach, which is necessary for the operationalization of the model. The discussion is mainly about data requirements in regard to travel demand; supply data are available within the existing four-step models and can be derived from a number of alternative data sources. A discussion of the methodological challenges and techniques related to the multistage development trajectory follows. Next, the modular framework that has been implemented to translate the methodological challenges to an operational platform is discussed. In the final section conclusions are drawn and topics for future development and research are discussed.

SPECIFIC ACTIVITY-BASED DATA COLLECTION METHODOLOGIES

Data for Modeling Dynamic Activity–Travel Behavior

The data requirements of the static and the dynamic model applications that have been outlined above constitute a real challenge. In particular, the dynamic activity–travel model application needs considerable additional effort in relation to data collection. Therefore, in addition to traditional activity–travel diaries, the model needs data on activity (re)scheduling decisions of individuals, data on household multiday activity scheduling, data on life trajectory events and how they affect activity–travel decisions, data on how individuals learn, and data on how short-term dynamics are linked to long-term decisions. Such data are available in typical cross-sectional travel surveys and time use surveys, and some need to be collected by means of a panel survey. In fact, in Flanders, there are no data available on either activity–travel schedules or panel surveys. The data collection therefore involved an extensive hybrid, multimethod approach. The various efforts are described below.

PARROTS Tool

The use of enhanced data collections [as reported in several application areas; see for instance Goulias and Janelle (*7*)] is particularly important in the dynamic application case because rescheduling decisions are probably not only undertaken at the level of activity, but are consequently probably also reflected in travel execution (e.g., other routes taken). Furthermore, automated data collection techniques are particularly well suited to obtain data that require a significant effort from the respondent, such as in the rescheduling of activities for the development of dynamic models.

For that purpose, an automated activity–travel diary survey tool called PARROTS [PDA (personal digital assistant) system for Activity Registration and Recording of Travel Scheduling] uses the Global Positioning System (GPS) to automatically record location data (*8*). The PDA was programmed such that its location is automatically registered, and respondents can provide information about their activity–travel behavior as well. Planned and executed activities and trips are registered with the possibility to alter all attributes of the planned activities. In that way, information is collected concerning the decision and scheduling processes, which results in an evolution from an intention to execute some activities and trips to an executed activity–travel diary. A similar philosophy was adopted in Rindsfüser et al. (*9*). Replanning information in the present case, however, is collected by allowing the respondent to update all attributes and by querying the reasons for the registered changes.

Currently, about 900 persons have been questioned by means of the PARROTS tool, which means that this study is probably one of the largest using GPS in the field of activity–travel data collection and one of the few that the authors are aware of that uses GPS-enabled PDAs. Also the weekly survey period makes it unique in the field. Bellemans et al. provide more detailed analyses with respect to the collected data by means of PARROTS, such as analysis of the impact of GPS-enabled PDA technology on user response rates, the impact of PDA technology on the quality of the collected diary data, and PARROTS usage patterns (*8*). The functional design of the tool has been discussed in Kochan et al. (*10, 11*).

Paper-and-Pencil Data Collection

Another part of the sample (about 1,500 persons; part of them belonging to the same households as the people who are questioned by means of the PARROTS tool, therefore enabling future modeling of intrahousehold decision making) is being questioned by means of a traditional paper-and-pencil method to account for the sample bias introduced when only computer-assisted forms of data collection are used. Furthermore, this choice enables the carrying out of comparative studies with respect to the behavior of both target groups in regard to response rates, experience, and so forth.

Social Network Data

At the time the PDA is picked up, participants are also questioned about their social network. This takes place during a short interview using Wellmann's instrument (*12*). In the application of this method, one obtains information about egocentric social networks by using only one name generator per group. Questions were asked about people the respondent feels closest to; these could be friends, neighbors, or relatives. The named alters were recorded and described in detail for parents, brothers and sisters, other family members, friends, neighbors, colleagues, and (sports-)club members.

Stated Adaptation Experiments

As mentioned previously, the goal of collecting data that measure the adaptation behavior of people aims at developing a dynamic component that more efficiently captures the complex process of activity generation and therefore enhances the behavioral realism of activity-scheduling models. However, decisions that constitute the short-term adaptation process of people are not trivial, to be captured solely by means of activity–travel diaries (e.g., activities that have been undertaken more than a week previously). For that reason, and for benchmarking purposes (with the weekly activity-diary information that has been collected), a specific Internet-based stated preference experiment was undertaken to gather additional data. Although this technique can

be used for different activities, it proved to be relevant particularly for flexible nonroutine activities that are frequently scheduled. More detailed information about the analysis results of the collected data can be found in van Bladel et al. (*13, 14*).

Event-Based (Long-Term) Data Collection

Finally, a dynamic model should ideally also account for a continuous change over time as a function of life trajectory events. An ideal scenario, of course, would be to keep the sample for an entire year and refresh it thereafter, thereby measuring people's activity and travel behavior at each of the different events (ideally before, during, and after the occurrence of the key event). The solution for capturing this type of travel information in relation to regular and key events was to implement a long-term panel survey that has been carried out by means of a Vehicle Embedded Data acquisition Enabling Tracking and Tracing (VEDETT) device, which has been specifically developed for this purpose. The logged in-vehicle data of the VEDETT tool can be transmitted to a central data collection point as a real-time data stream or in batches. The system was installed and is currently running in 14 vehicles. A website application has been developed for the survey participants to communicate with the VEDETT device. On the website, the motivations or reasons behind all the trips made by car and all the additional travel facets can be indicated. To minimize the burden for the participators in the long-term field trial, addresses that are frequently visited can be designated as points of interest (POIs). Second, the system is embedded with a self-learning capacity, to allow for some trips from two POIs that are frequently made, to automatically suggest the motivation and number of passengers. More detailed information about the VEDETT application can be found in Broekx et al. (*15*).

STATIC AND DYNAMIC ACTIVITY-BASED MODELING

As mentioned previously the first step in the four-stage development trajectory of the model includes the development of a static activity-based model for Flanders. Although the emphasis in this paper is clearly on developing a framework allowing for the development of dynamic activity-based models, it has been highlighted that advancing directly toward a full agent-based microsimulation framework is not always the most appealing prospect from an end user point of view. The rest of this section describes the current status and future research steps in the development of the four-stage trajectory.

Static Activity-Based Modeling

The scheduling model currently implemented in the FEATHERS framework is based on the scheduling model present in Albatross (*4*). Currently, the framework is fully operational at the level of Flanders. The real-life representation of Flanders is embedded in an agent-based simulation model that consists of more than 6 million agents, each agent representing one member of the Flemish population. The scheduling is static and based on decision trees, in which a sequence of 27 decision trees is used in the scheduling process. Decisions are made on the basis of a number of attributes of the individual (e.g., age, gender), of the household (e.g., number of cars), and of the geographic zone (e.g., population density, number of shops). For each agent with his or her specific attributes, it is, for example, decided whether an activity is performed or not. Subsequently, the location, transport mode, and duration of the activity, among other aspects, are determined, taking into account the attributes of the individual.

The Albatross model was reimplemented in the FEATHERS framework for the Flanders study area (see the following section for a more detailed description). Because of the modular framework, the model can be rapidly adapted to use other core scheduling models relying on other artificial intelligence techniques such as Bayesian networks (*16, 17*), simple classifiers (*18*), and association rules (*19*). Thus, in addition to the specific tree induction algorithm used in Albatross, users can opt for a wide variety of knowledge extraction techniques and meanwhile benefit from the functionalities that are provided by the other FEATHERS modules (e.g., preprocessing). Ease of transferability of the models to other study areas, which has been investigated for a static activity-based model in, for example, Arentze et al. (*20*), was a design goal of FEATHERS. This resulted in provisions for easy incorporation of the input data of new study areas into the system and in an easy extension of existing data sets with new, context-specific attributes.

Within the FEATHERS framework, the developed activity-based models are microsimulation models, simulating each member of the population individually. Hence, for each of the car trips generated, one can obtain information on the type of activity associated with this trip as well as other context information such as, for example, socioeconomic data on the traveler and the traveler's activity schedule for the day. It is reasonable to assume that on an individual level, the context of a trip plays a role in how an individual assesses the cost/utility of a certain route, giving for instance preference to other routes taken in the context of flexible activities such as leisure. Preliminary research results indeed report on the significant impact of travel purpose on driving behavior, which illustrates the relevance of the trip context when behavior during trips is investigated (*21*).

The link between activity-based models and traffic assignment is a key factor in increasing the deployment of activity-based models in practice because the resulting visualization and network functionalities meet the needs and concerns of practitioners. Indeed, the traditional network assignment functionality has always existed before in four-step models. Hence, in this first stage, the link between activity-based models and traffic assignment results in a coupling of new activity-based modeling techniques with models and applications that have been operational in practice for a long time.

Semistatic Activity-Based Modeling

Because of the microsimulation of activity–travel patterns, most activity-based models do not suffer from aggregation biases. Microsimulation provides a practical method with which to implement probabilistic models at the level of the individual. The basic argument is that people travel, not zones, and by averaging to the level of zones, much information is lost and the aggregation bias is significant. Because of microsimulation it is possible to produce, for instance, origin–destination (O-D) matrices at an hourly (or even more detailed) level, for different days in the week (see the section on specific activity-based data collection methodologies for data requirements), or under specific circumstances such as extreme weather conditions. However, the behavioral modeling process in itself is not changed.

Indeed, it is known that most currently operational activity-based models are applicable only in a stationary environment. This charac-

teristic is inconsistent with other studies in which it has been proved that travel behavior is highly evolutionary and nonstationary (*22*).

To that end, some initial studies have been undertaken to extract nonstationary information from longitudinal data. In an initial application, traffic counts have been used to observe the impact of day of the week, but also the impact of regular events, such as holidays, on the observed traffic states (*23, 24*). Weather information has also been accounted for. The different techniques pointed out the significance of the day-of-the-week effects: weekly cycles seem to determine the variation of daily traffic flows. With respect to weather information, the most appealing result for policy makers is the heterogeneity of the weather effects among different traffic count locations. Furthermore, the results indicated that precipitation, cloudiness, and wind speed have a clear diminishing effect on traffic intensity, whereas maximum temperature, sunshine duration, and hail significantly increase traffic intensity.

Obviously, these analyses are only preliminary. Tools such as the VEDETT application discussed previously further allow for a more detailed behavioral impact study, enabling one to keep the sample for an entire year, thereby measuring and comparing people's (nonstationary) activity and travel behavior before, during, and after the occurrence of an event. The functionalities required to accommodate the data for the analyses discussed above are currently operational in FEATHERS.

Dynamic Activity-Based Model

The next step in the trajectory deals with the development of a dynamic agent-based microsimulator that allows one to simulate activity–travel scheduling decisions, within day rescheduling, and learning processes in high resolution of space and time. A priori, the dynamic activity-based simulation system is based on the Aurora framework, a full dynamic activity-based model focusing on the rescheduling of activity–travel patterns.

The basis of the Aurora implemented model appear in Timmermans et al. (*25*) and Joh et al. (*26, 27*) focusing on the formulation of a comprehensive theory and model of activity rescheduling and reprogramming decisions as a function of time pressure. Apart from duration adjustment processes, Aurora also incorporates other potential dynamics such as change of destination, transport mode, and other facets of activity–travel patterns. Later, this model was extended to deal with uncertainty (*28*), various types of learning (*29*), and responses to information provision (*30, 31*). Finally, the model has been implemented as a multiagent simulation system (*32*). Currently, some proof of concepts for this third stage in the deployment process are operational in FEATHERS.

Full Microscopic Activity-Based Model with Microscopic Route Choice

Given the level of detail of the activity-based models discussed in the previous sections, the implementation of the bidirectional interaction between the activity-based model and the transportation system on a nonmicroscopic level exhibits some drawbacks.

The O-D matrices that are constructed on the basis of the predicted activity–travel diaries can be aggregated at different levels of detail. Although it is desirable to retain as much information as possible and hence work at a low level of aggregation, the level of disaggregation of the O-D matrices is quite limited in practice, for example, a matrix segmentation by trip purpose only. Although some other general sociodemographic variables can additionally be accounted for in the segmentation, the assignment procedures that are used in the conventional four-step models and in Stage 1 remain limited in the maximum level of disaggregation of the matrix that can be dealt with.

The presence of uncertainty and of incomplete information can yield a discrepancy between the attributes of intended and executed activities or trips. This issue is dealt with by dynamic activity-based models by introducing the concept of schedule execution as presented in the previous section. This schedule execution introduces a feedback between the state of the transportation network and the scheduling process. By using nonmicroscopic traffic assignment algorithms, the agent-based concept is broken and the concept of individual route choice is replaced by a model of a higher level of aggregation. This aggregation restricts the level of detail at which effects of policies on the behavior of (very specific groups of) individuals can be assessed.

The issues discussed above are resolved by incorporating microscopic route choice behavior in the dynamic activity-based model. Individual travelers in this case are endowed with the capability to consider alternatives with respect to their intended route, enabling them to cope with changes in the traffic state in an autonomous manner. Indeed, traffic assignment is inherently dynamic in the sense that the traffic state of the road network changes frequently. Consequently, the optimal route of a traveler can be affected by changes in the traffic state. Such changes typically lead to travelers reassessing their current situation and considering alternative routes. However, changes in the traffic state introduce rerouting behavior, and because of the schedule execution mechanism, information on the traffic state of the transportation network effectively propagates toward the agent-based scheduling process. In this way, schedules that are consistent with the traffic state on the transportation network can be achieved. Enabling microscopic route choice within the FEATHERS framework is the topic of ongoing activities.

FEATHERS MODULAR SYSTEM DESIGN

Facing the challenge to be able to implement several new theoretical advances such as those that are reflected in the four-stage development process in the FEATHERS platform, a modular framework to conduct research on agent- and activity-based models has been developed. The modularity of the FEATHERS framework is guaranteed by means of the module-based design and by the use of the object-oriented paradigm. This design results in an agile environment that allows for easy removal, exchange, and insertion of functionalities and even complete modules.

An overview of the current modular structure of the FEATHERS framework is presented in Figure 1. In the rest of this section, the functionality of the modules and the implications of the four-stage timeline on the evolving functionality of the modules will be discussed.

Configuration Module

To be able to exploit FEATHERS' modular structure to the maximum extent, a flexible configuration functionality is required. Every module that is active in FEATHERS communicates with the configuration module (ConfMod) to obtain its specific required settings (see Figure 1). This approach allows for a central configuration management, from which the relevant settings are dispatched to each of the

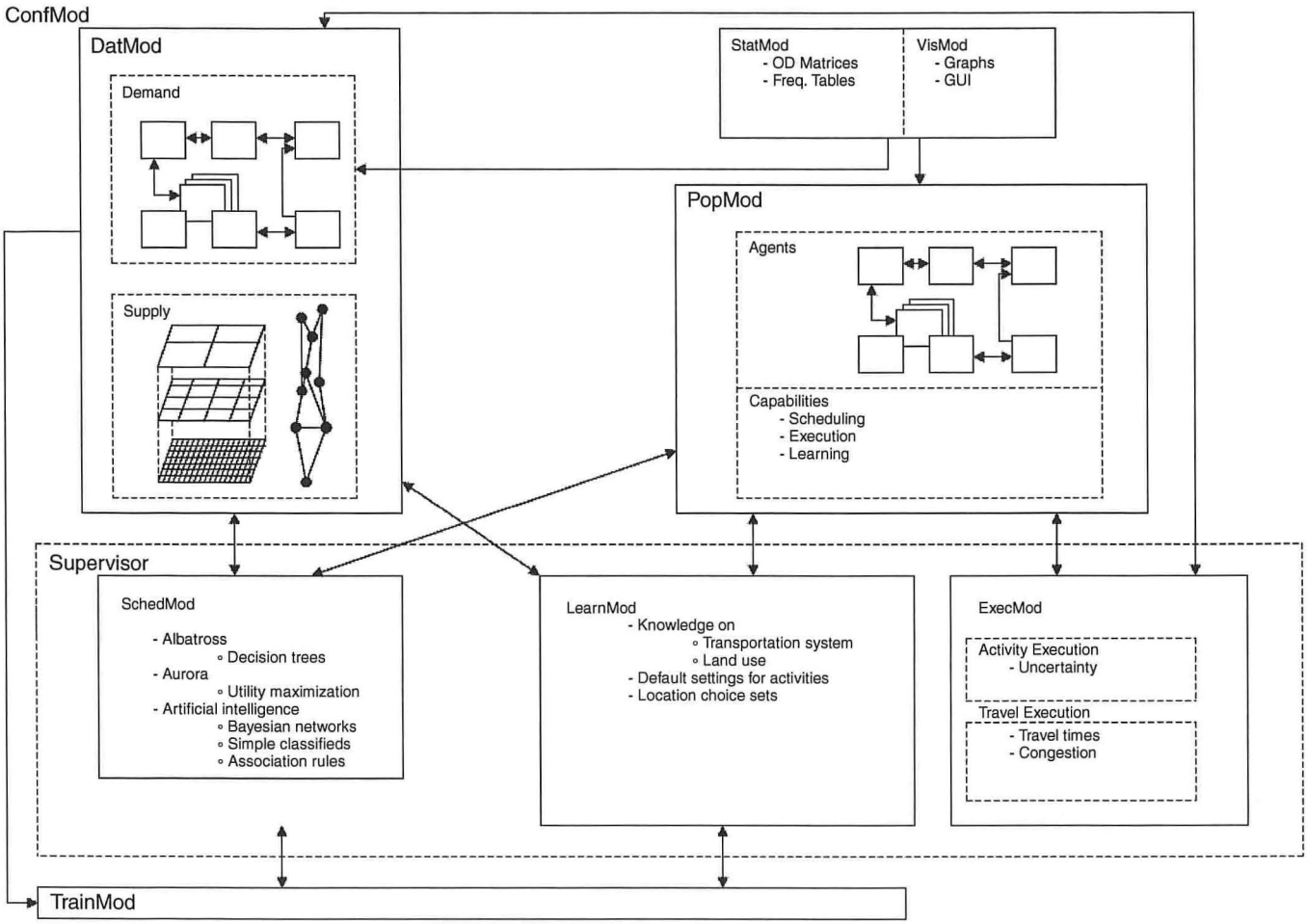


FIGURE 1 Schematic overview of FEATHERS modules and their functionalities and interactions.

modules. Modules can be switched (in-)active by using the ConfMod to facilitate the multistage development strategy described above. If for a module no settings are available in the configuration file, it is considered to be inactive by default. This way users are not burdened by functionality that is provided by the framework but that is not needed for the current experiments (compare simultaneous development of functionalities for several stages).

To guarantee extensible and structured configuration settings, which are required to accommodate future and currently unknown configuration settings, the configuration module stores all the configuration settings for the FEATHERS modules in XML format (*33*). This makes the addition of new parameter settings for a (new) module a simple matter of updating the XML configuration file.

Data Module

One of the core modules in the system is the data module (DatMod). It provides access to the data that need to be accessible throughout all other modules. Two major types of data are provided by the DatMod: supply and demand data (see Figure 1).

The (geographic) supply data include not only the transportation network but also information on geographic zones in the study area, such as the attractiveness of a zone for conducting certain activities. Also information on the availability and performance of the transportation system between the zones in the study area (e.g., travel times, travel costs, bus fares) is included in the geographic supply data. In summary, the supply data consist of the data describing the context in which the agents live and schedule their activity and travel episodes.

The demand data (see the upper part of the DatMod block in Figure 1) consist of the activity–travel diaries or schedules that describe the demand for the execution of activities at certain locations as well as the resulting demand for transportation. The collected diaries are typically accompanied by person and household data for the persons executing the diaries. The data model for the demand data in the FEATHERS DatMod is aware of the following entities: persons, households, (optionally) cars, activities, journeys, and lags and assumes they relate as presented in Figure 2. Because FEATHERS is not tailored only toward the Flemish situation and the data survey discussed previously, the attributes that are available in the data files for each of the entity types are fully customizable through the ConfMod.

The supply and the demand data managed by the DatMod are made available to other modules through the DatMod's standardized interface.

Because it is imperative that the demand data be easily accessed by (future) modules, it is important to efficiently implement the relationships between the entities in the data model. These relationships are defined in the data model presented in Figure 2. Because the number of persons and households in a survey is typically rather small

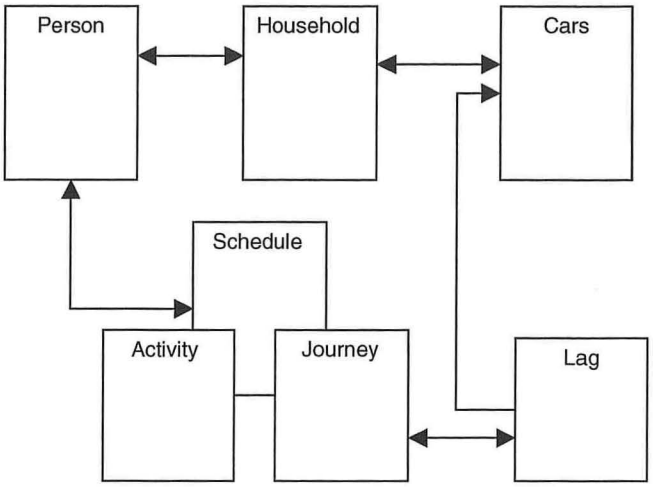


FIGURE 2 Schematic representation of relations between transportation demand data entities in FEATHERS data module.

(e.g., 2,500 households for the survey discussed in this paper), the demand data can be loaded into memory for fast access.

Because not all geographic supply data are available at the same level of detail, the DatMod provides support for different levels of detail (currently three; expandable if required). This support includes keeping track of the relation between the zones at the different levels of detail. In the current implementation it is assumed that each zone at the lower level (more detail) belongs to one higher level zone (less detail) only. These relations between the levels of geographic detail allow for (dis-)aggregation of simulation results to the desired geographic level of detail.

The attributes that are stored for the zones in a supply data layer are configured through the ConfMod for flexibility. For the Flanders study area (total area of approximately 13,500 km^2) the levels of detail used are statistical sector (small administrative unit, comparable to districts or quarters, 10,255 zones), submunicipalities (1,145 zones), and municipalities (327 zones). Because the number of zones in each of the geographic data layers is rather limited for the study area, it is perfectly feasible to load all data in memory for fast access. Although it was not required for the current research, a configuration setting allows the DatMod to switch over from loading all data into memory to using direct access binary data files if sufficient memory is not available. This switch is transparent to the modules consulting the data.

Because information on the transportation system (e.g., bus fares between zones) cannot be attributed to one zone only, the DatMod also provides attributes for pairs of zones for each of the levels of geographic detail. The attributes that are stored for each pair of zones are configured through the ConfMod. However, as the required storage capacity increases with the square of the number of zones, the DatMod provides the choice between loading all data in memory and using direct access files. For the Flemish case study, the data on pairs of municipalities and on submunicipalities were loaded into memory, whereas for the statistical sectors a direct access file was used.

The supply data on the attractiveness of zones for the execution of activities used for the model in Flanders are exceptionally rich because of the availability of the socioeconomic survey, in which the survey of the full Flemish population (6 million) contained obligatory questions on several sociodemographic variables (age, gender, etc.). In addition to sociodemographic variables, the data set contains commuting behavior of all persons in the study area (population level). Given this characteristic, one can derive from these data, for example, the level of employment by employment sector for each statistical sector, which can be used to calculate the availability and attractiveness of locations for different activities. Information about the transport system (road network data, congested travel times, etc.) is available from the existing four-step model currently used in Flanders. Also the traffic network that is used (see Figure 1) results from the existing four-step model managed by the Flemish government. Although the DatMod manages geographic data, it currently does not provide geographic information system (GIS) functionalities. Hence, geographical manipulations such as, for example, overlays and map matching of GPS data need to be performed in a preprocessing step and the resulting data need to be imported into the FEATHERS DatMod afterward.

Population Module

The units of investigation in an activity-based model are the persons making scheduling decisions that result in activity–travel diaries. Hence, the agents in an agent-based, activity-based model are the individual persons. During scheduling, the agent's person characteristics or attributes are used as inputs for the scheduler to drive the simulated decisions of the agent. The definition of which attributes of the agents are used is realized through the ConfMod. Examples of commonly used person attributes are marital status, age, and possession of a driver's license.

Similar to the person entities in the DatMod, the persons (agents) in the population module (PopMod) relate to households, car (optional), activity, journey and lag entities (Figure 2). In the PopMod, these entities are virtual entities as opposed to the real entities in the DatMod. Through the relations between the entities, the attributes of all entities are accessible to be used in the agent's scheduling process in addition to the person attributes.

An important difference between the person entities in the DatMod and the agents in the PopMod is the fact that the agent entities possess important additional functionalities: scheduling, schedule execution, and learning (Figure 1), which are implemented in the scheduling module, the schedule (activity and travel) execution module, and the learning module, respectively. These functionalities are implemented in separate modules to make replacement and extension of agent functionalities as convenient as possible.

To perform a simulation of activity and travel behavior of individuals in a population, a synthetic population consisting of persons and households (and optionally cars belonging to the household) needs to be built. The PopMod is responsible for the management of the different agents (persons) that are used in the synthetic population. The synthetic population therefore consists of a collection of agents in which each agent is characterized by a number of attributes. As mentioned previously, the data required are available at the population level in Flanders by means of the socioeconomic survey. These population data can then be updated to the current prediction year by the use of the iterative proportion fitting (IPF) technique. IPF is a well-established technique, with the theoretical and practical considerations behind the method thoroughly explored and reported in the literature (*34*). It uses the population or the larger sample margins to update the information at the cell frequency level. Several applications of the technique in travel demand modeling have been reported (*35–37*).

A common functionality of all agents throughout the four development stages is the scheduling functionality. On the basis of the agent's personal, household-related, environmental, and schedule-related attributes, the agent is able to predict an activity–travel schedule by using functionalities provided by the scheduling module. The resulting activity and travel episodes for an agent are stored in the activity, journey, and lag entities linked to that agent (Figure 2). During the simulation, the person, household and, optionally, car entities of the agents (corresponding to the upper part of Figure 2) are used to predict the schedules for the agents, which constitute an important model output and which correspond to the lower part of Figure 2.

Schedule Module

The schedule module (SchedMod) is a generic module in which different scheduling algorithms can be implemented. The ConfMod determines which of the available scheduling algorithms is activated. The SchedMod is tightly interfaced with the (agents in the) population module because it implements the scheduling algorithm that uses input data from the PopMod and stores the results in the schedules in the PopMod.

In the scope of Stages 1 and 2 of the FEATHERS development trajectory, a decision tree–based scheduling algorithm was implemented in the schedule module. This implementation currently consists of a sequence of 27 decision trees, in which each decision tree is used to model decisions on specific activity–travel schedule properties (e.g., going to work or not, transport mode for a journey, start time and duration of an activity). Besides the decision trees, the scheduling mechanism contains an algorithm to make the schedules consistent. To be consistent a schedule needs to comply with a number of constraints: situational constraints (one cannot be in two places at the same time), institutional constraints (opening hours constrain certain activity behavior), household constraints (bringing children to school), spatial constraints (particular activities cannot be performed at particular locations), time constraints (activities require some minimum duration), and spatial–temporal constraints (travel time depends on transport mode). The output of the scheduler in the SchedMod is the collection of activity–travel diaries for all agents in the PopMod.

Another, more advanced scheduler that is being investigated with the FEATHERS framework uses the diary utility maximizing approach. Although this scheduling approach is fundamentally different from the decision tree–based scheduler, both schedulers were implemented in the scheduling module side by side. This illustrates the flexibility of the design of the schedule module in the framework. This flexibility enables further research on alternative innovative scheduling mechanisms and allows for benchmarking of the schedulers (*38*).

All stages of the four-stage development trajectory discussed in this paper require all of the modules discussed above to be operational.

Schedule Execution Module

A dynamic activity-based model as described in Stages 3 and 4 requires a schedule execution mechanism. This schedule execution mechanism is implemented in the schedule execution module (ExecMod) of the FEATHERS framework (see Figure 1) and simulates the simultaneous and synchronous execution of all activities and journeys for all agents. As can be observed from Figure 1, separate modules are provided for simulating the execution of activities and for travel execution.

In the activity execution module, uncertainty on the scheduled activities can be modeled. Indeed, during the execution of activities unforeseen events can take place resulting in changes of activity attributes, for example, the duration of the activity, compared with the attributes of the activity as it was originally scheduled.

In the travel execution module the relation between traffic demand and the performance of the transportation system (e.g., car travel speeds on a link as a function of traffic intensity) is accounted for. Because the agent's schedule executions are simulated for all agents simultaneously, the total traffic demand can be computed for each transportation mode and at each moment in time. To obtain the traffic intensities on links in the transportation network, the traffic demand needs to be loaded onto the network. In Stage 4 this is achieved by simulating the microscopic route choice behavior.

The potential mismatch between the attributes of scheduled and simulated executed activities or travel results in a potential inconsistency in the schedules if no corrective rescheduling action is taken.

The rescheduling functionality, combined with the traffic assignment from the Stage 4 model, results in a bidirectional coupling between the scheduling and the transportation network: the traffic demands predicted by the activity-based model affect the traffic states in the transportation network and vice versa.

Rescheduling of activities and travel is managed in the FEATHERS framework by the supervisor (see Figure 1), which coordinates between the scheduling and the schedule execution for each agent. This coordination consists mainly of deciding when to check the partially executed schedule for inconsistencies and when to start the rescheduling (SchedMod) and the schedule execution (ExecMod).

Learning Module

The learning behavior of persons stems from the fact that they observe that their assumed knowledge about the environment in which they operate (e.g., the transportation network) does not match reality. An indication of this mismatch is given by a mismatch between scheduled and executed activities or travel. The learning process of the agents is managed by the supervisor in combination with the (re-)scheduling and the schedule execution for that agent. The supervisor takes into account that the rescheduling processes typically run on a faster time scale than the learning processes. By adaptation of the supervisor and the scheduling, schedule execution, and learning modules (LearnMods), a wide range of experiments can be conducted.

Statistics and Visualization Modules

The statistics module (StatMod) provides reports concerning the (synthetic) population and the activity–travel schedules to the FEATHERS user. This includes information that can be extracted at the level of households (e.g., distribution of households according to availability of means of transportation); persons (e.g., use of transportation modes), journeys (e.g., average number of journeys per day), lags (e.g., average number of lags per journey), and activities. Given the similarity in the person, household, car, activity, journey, and lag entities and their relations in the DatMod and the PopMod, the StatMod and the visualization module (VisMod) make abstraction from the fact whether they consult the DatMod or the PopMod

to extract the data to report to the user. Hence, statistics that are implemented for the survey data in the DatMod can readily be used to draw the corresponding statistics on simulated data from the PopMod. Which statistics are to be drawn by the StatMod is configured through the ConfMod.

Because the activity–travel diaries contain detailed travel information, the StatMod provides the functionality of skimming through all schedules and compiling an O-D matrix. Given the level of detail of the data, the travel information can be aggregated in segmented O-D matrices such as time-sliced O-D matrices, O-D matrices per transportation mode, and O-D matrices per activity type. This functionality enables a transition step in the evolution from four-step models toward activity-based models by exporting O-D matrices that are assigned to the transportation network by using the traffic assignment tools from the traditional four-step model as was discussed in Stage 1.

The VisMod relates strongly to the StatMod in the sense that the VisMod will create graphic reports contrary to the numerical reports provided by the StatMod. Currently, the VisMod is not yet operational and all FEATHERS reports are obtained through the StatMod. However, to improve user friendliness, a graphical user interface and a VisMod will be added to the FEATHERS framework in the future.

Training Module

All models used throughout the FEATHERS framework need to be calibrated by using real-life data. This functionality is provided by the training module (TrainMod). The TrainMod is configured through the ConfMod and obtains the required data from the DatMod. The output of the TrainMod is calibrated model parameters for the models that are used in the other modules (see Figure 1).

CONCLUSIONS

The main goal of the FEATHERS framework, which has been presented in this paper, is to allow for easy updating and/or replacement of functionalities used in activity-based models as the state-of-the-art in the activity-based research field progresses rapidly. It is therefore believed that the modular framework holds considerable promise to facilitate the research on and the development of dynamic activity-based models for transport demand.

It was illustrated that the modular design of the FEATHERS framework is compatible with a long-term, four-stage development trajectory of activity-based models that was postulated for Flanders (Belgium): Stage 1 is the development of a static activity-based model; Stage 2 is the development of a semistatic model accounting for evolutionary and nonstationary behavior; Stage 3 is the development of a fully dynamic activity-based model including short-term adaptation (rescheduling) and learning; and Stage 4 is a full agent-based dynamic activity-based microsimulation framework including traffic assignment. Besides the discussion of the different modules in the FEATHERS framework and their interactions, it was shown how the FEATHERS modules' functionalities accommodate the requirements of each of the four development stages.

It has been shown that data collection is a prerequisite for the application of static and dynamic activity-based models. To that end, an extensive hybrid, multimethod data collection approach has been described in detail. It was shown that in particular the dynamic activity–travel model application needs considerable additional effort in regard to data collection. It has been shown that in addition to traditional activity–travel diaries, such a model needs data on activity rescheduling decisions of individuals, data on household multiday activity scheduling, data on life trajectory events and how they affect activity–travel decisions, data on how individuals learn, and data on how short-term dynamics are linked to long-term decisions.

REFERENCES

1. Bhat, C. R., J. Y. Guo, S. Srinivasan, and A. Sivakumar. A Comprehensive Micro-Simulator for Daily Activity-Travel Patterns. *Proc., Conference on Progress in Activity-Based Models,* Maastricht, Netherlands, May 2004, pp. 28–32.
2. Pendyala, R. M., R. Kitamura, A. Kikuchi, T. Yamamoto, and S. Fujji. FAMOS: Florida Activity Mobility Simulator. Presented at 84th Annual Meeting of the Transportation Research Board, Washington, D.C., 2005.
3. Arentze, T. A., and H. J. P. Timmermans. *Albatross: A Learning-Based Transportation Oriented Simulation System.* European Institute of Retailing and Services Studies, Eindhoven, Netherlands, 2000.
4. Arentze, T. A., and H. J. P. Timmermans. *Albatross 2: A Learning-Based Transportation Oriented Simulation System.* European Institute of Retailing and Services Studies, Eindhoven, Netherlands, 2005.
5. Davidson, W., R. Donnelly, P. Vovsha, J. Freedman, S. Ruegg, J. Hicks, J. Castiglione, and R. Picado. Synthesis of First Practices and Operational Research Approaches in Activity-Based Travel Demand Modeling. *Transportation Research Part A,* No. 41, No. 15, 2007, pp. 464–488.
6. Balmer, M., K. Nagel, and B. Raney. Agent-Based Demand Modelling Framework for Large Scale Micro-Simulations. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1985,* Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 125–134.
7. Goulias, K., and D. Janelle. *GPS Tracking and Time-Geography: Applications for Activity Modeling and Microsimulation.* Final report. FHWA-sponsored Peer Exchange and CSISS Specialist Meeting, 2006.
8. Bellemans, T., B. Kochan, D. Janssens, G. Wets, and H. J. P. Timmermans. Field Evaluation of Personal Digital Assistant Enabled by Global Positioning System: Impact on Quality of Activity and Diary Data. In *Transportation Research Record: Journal of the Transportation Research Board No. 2049,* Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 136–143.
9. Rindsfüser, G., H. Mühlmans, S. T. Doherty, and K. J. Beckmann. Tracing the Planning and Execution of and Their Attributes—Design and Application of a Hand-Held Scheduling Process Survey. Presented at 10th International Conference on Travel Behaviour Research, Lucerne, Switzerland, Aug. 10–15, 2003.
10. Kochan, B., T. Bellemans, D. Janssens, and G. Wets. Dynamic Activity-Travel Diary Data Collection Using a GPS-Enabled Personal Digital Assistant. *Proc., Innovations in Travel Demand Modeling Conference,* Austin, Tex., May 21–23, 2006.
11. Kochan, B., T. Bellemans, D. Janssens, and G. Wets. Dynamic Activity-Travel Diary Data Collection Using a GPS-Enabled Personal Digital Assistant. Presented at 9th International Conference on Applications of Advanced Technology in Transportation, Chicago, Ill., Aug. 13–16, 2006.
12. Wellmann, B. The Community Question: The Intimate Networks of East Yorkers. *The American Journal of Sociology,* Vol. 84, No. 5, 1979, pp. 1201–1231.
13. van Bladel, K., T. Bellemans, G. Wets, T. A. Arentze, and H. J. P. Timmermans. Fitting S-Shaped Activity Utility Functions Based on Stated-Preference Data. Presented at 11th International Conference on Travel Behaviour Research, Kyoto, Japan, Aug. 16–20, 2006.
14. van Bladel, K., T. Bellemans, D. Janssens, G. Wets, L. Nijland, T. A. Arentze, and H. J. P. Timmermans. Design of Stated Adaptation Experiments: Discussion of Some Issues and Experiences. Presented at International Conference on Survey Methods in Transport: Harmonization and Data Comparability, Annecy, France, May 25–31, 2008.
15. Broekx, S., T. Denys, and L. Int. Panis. Long-Term Travel Surveys: How Can the Burden Remain Bearable. Presented at 33rd Colloquium

Vervoersplanologisch Speurwerk, Amsterdam, Netherlands, Nov. 23–24, 2006.
16. Janssens, D., G. Wets, T. Brijs, K. Vanhoof, T. A. Arentze, and H. Timmermans. Improving Performance of a Multiagent Rule-Based Model for Activity Pattern Decisions with Bayesian Networks. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1894,* Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 75–83.
17. Janssens, D., G. Wets, T. Brijs, K. Vanhoof, T. A. Arentze, and H. J. P. Timmermans. Integrating Bayesian Networks and Decision Trees in a Sequential Rule-Based Transportation Model. *European Journal of Operational Research,* Vol. 175, No. 1, 2006, pp. 16–34.
18. Moons, E. *Modelling Activity-Diary Data: Complexity or Parsimony?* PhD dissertation. Limburg University, Diepenbeek, Belgium, 2005.
19. Keuleers, B., G. Wets, T. Arentze, and H. Timmermans. Association Rules in Identification of Spatial-Temporal Patterns in Multiday Activity Diary Data. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1752,* TRB, National Research Council, Washington, D.C., 2001, pp. 32–37.
20. Arentze, T., F. Hofman, H. van Mourik, and H. Timmermans. The Spatial Transferability of the Albatross Model System: Empirical Evidence from Two Case Studies. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1805,* Transportation Research Board of the National Academies, Washington, D.C., Jan. 2002, pp. 1–7.
21. Beckx, C., L. Panis, G. Wets, R. Torfs, C. Mensink, S. Broekx, and D. Janssens. Impact of Trip Purpose on Driving Behaviour: Case Study on Commuter Behaviour in Belgium. *Proc., of 15th International Symposium on Transport and Air Pollution,* Reims, France, Vol. 15, No. 2, 2006, pp. 332–337.
22. Schönfelder, S. *Urban Rhythms: Modelling the Rhythms of Individual Travel Behaviour.* PhD dissertation. Eidgenössische Technische Hochschule, Zürich, Switzerland, 2006.
23. Cools, M., E. A. Moons, and G. Wets. Investigating Effect of Holidays on Daily Traffic Counts: Time Series Approach. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2019,* Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 22–31.
24. Cools, M., E. A. Moons, and G. Wets. Assessing the Impact of Weather on Traffic Intensity. Presented at 87th Annual Meeting of the Transportation Research Board, Washington, D.C., 2008.
25. Timmermans, H. J. P., T. A. Arentze, and C.-H. Joh. Modelling Effects of Anticipated Time Pressure on Execution of Activity Programs. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1752,* TRB, National Research Council, Washington, D.C., 2001, pp. 8–15.
26. Joh, C.-H., T. A. Arentze, and H. J. P. Timmermans. Understanding Activity Scheduling and Rescheduling Behaviour: Theory and Numerical Simulation. In *Modelling Geographical Systems* (B. Boots et al., eds.), Kluwer Academic Publishers, Dordrecht, Netherlands, 2003, pp. 73–95.
27. Joh, C.-H., T. Arentze, and H. Timmermans. Activity-Travel Scheduling and Rescheduling Decision Processes: Empirical Estimation of Aurora Model. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1898,* Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 10–18.
28. Arentze, T. A., and H. J. P. Timmermans. A Theoretical Framework for Modeling Activity-Travel Scheduling Decisions in Non-Stationary Environments Under Conditions of Uncertainty and Learning. *Proc., International Conference on Activity-Based Analysis* (CD-ROM), Maastricht, Netherlands, 2004, pp. 13.
29. Arentze, T. A., C. Pelizaro, and H. J. P. Timmermans. Implementation of a Model of Dynamic Activity-Travel Rescheduling Decisions: An Agent-Based Micro-Simulation Framework. *Proc., Computers in Urban Planning and Urban Management Conference* (CD-ROM), London, 2005, pp. 16.
30. Arentze, T. A., and H. J. P. Timmermans. Information Gain, Novelty Seeking and Travel: A Model of Dynamic Activity-Travel Behavior Under Conditions of Uncertainty. *Transportation Research A,* Vol. 39, No. 2-3, 2005, pp. 125–145.
31. Sun, Z., T. Arentze, and H. J. P. Timmermans. Modelling the Impact of Travel Information on Activity-Travel Rescheduling Decisions Under Conditions of Travel Time Uncertainty. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1926,* Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 79–87.
32. Arentze, T. A., and H. J. P. Timmermans. A Cognitive Agent-Based Simulation Framework for Dynamic Activity-Travel Scheduling Decisions. *Proc., Knowledge, Planning and Integrated Spatial Analysis,* LISTA, 2005.
33. W3C (2006) eXtensible Markup Language (XML). World Wide Web Consortium (W3C). http://www.w3.org/XML.22.
34. Beckman, R. J., K. Baggerly, and M. McKay. Creating Synthetic Baseline Populations. *Transportation Research Part A,* Vol. 30, No. 6, 1996, pp. 415–429.
35. Arentze, T., H. J. P. Timmermans, and F. Hofman. Creating Synthetic Household Populations: Problems and Approach. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2014,* Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 85–91.
36. Guo, J. Y., and C. R. Bhat. Population Synthesis for Microsimulating Travel Behavior. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2014,* Transportation Research Board of the National Academies, 2007, pp. 92–101.
37. Wong, D. W. S. The Reliability of Using the Iterative Proportional Fitting Procedure. *Professional Geographer,* Vol. 44, No. 3, 1992, pp. 340–348.
38. Vanhulsel, M., D. Janssens, and G. Wets. Calibrating a New Reinforcement Learning Mechanism for Modeling Dynamic Activity-Travel Behavior and Key Events. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C., 2007.

The Transportation Demand Forecasting Committee peer-reviewed this paper.

Multiple Objectives in Travel Demand Modeling

Yaron Hollander

Traditional techniques for estimating travel demand models cannot always identify a model if the quality of the input data is poor. These techniques do not allow modelers to easily predefine types of travel behaviors that they or their clients believe cannot be true. Models estimated with the best academic practice also may occasionally fail important validation tests. These factors often lead practitioners to determine model parameters through an inefficient trial-and-error process. A multiobjective model estimation procedure is presented that overrules solutions that cannot meet either statistical or political criteria. This procedure is not intended to criticize the traditional modeling approach, but it illustrates that a more pragmatic approach is available and works efficiently. This conclusion is illustrated in the estimation of a demand model for Dublin, Ireland.

The decision to invest large amounts of money in transport infrastructure is informed by estimates of costs and benefits, and often these estimates are based on travel demand forecasts. It is widely agreed that the ability to choose sensibly which schemes and policies to promote depends on the availability of powerful demand forecasting tools.

The past few decades have seen the parallel evolution of different strands of demand modeling practice. The academic community has developed robust techniques for model estimation, paying various degrees of attention to practicalities such as limited data availability or constraints imposed externally on the modeling work. Simultaneously, practitioners and consultants have been using modeling approaches ranging from rigorous methods that comply with the best academic practice to pragmatic solutions with little theoretical support.

The demand modeling literature does not often provide solutions for cases in which the only available data are not of sufficient quality to use rigorous estimation techniques, or cases in which the modeling work is required to meet constraints specified by decision makers. The concept described in this paper was devised when the author failed to find in the literature a solution for a problem repeatedly experienced as a practitioner. A model estimation procedure is presented that is less rigorous than the techniques used widely but is applicable in a wider range of situations. One application of this procedure is illustrated in the estimation of a demand model for the Greater Dublin Area in Ireland.

The next section discusses the practical issues that made it necessary to create the modeling approach proposed here. The approach itself is described in the two following sections: first its main principles are explained and then the principles of the optimization technique it employs are reviewed. Subsequently, the Dublin case study is presented, followed by some conclusions.

Steer Davies Gleave, 28-32 Upper Ground, London, SE1 9PD, United Kingdom. Current affiliation: South Yorkshire Passenger Transport Executive, 11 Broad Street West, Sheffield S1 2BQ, United Kingdom. yarhol@gmail.com.

Transportation Research Record: Journal of the Transportation Research Board, No. 2175, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 120–129.
DOI: 10.3141/2175-14

MODEL ESTIMATION PRACTICALITIES

Typically travel demand modeling involves creating models to replicate the way travelers choose their time of travel, mode of travel, trip destination, and so forth. Models are often of a logit form, and the modeling work aims to determine the parameter values for their utility functions, some scaling parameters, or both. The input data come either from revealed preference (RP) sources (travel diaries, passenger surveys, roadside interviews, traffic counts) or from stated preference (SP) studies.

When models are estimated from SP data, the common practice is to determine the values of the model parameters by using a maximum likelihood approach (*1*). This is a powerful and robust technique, and the set of parameter values it renders as the solution of the estimation problem is often the global optimum. Different variants, such as maximum simulated likelihood, are used for variants of the logit model (*2*). Recently, Bayesian estimation has been identified as an alternative approach that appears similarly credible (*2*). In the academic community, few other estimation techniques are considered acceptable.

Maximum likelihood estimation works well with SP data because SP surveys are based on statistical designs that ensure that the data encompass a wide range of choice situations. This makes the constraining and identification of parameters relatively convenient. Rich sources of RP data that exhibit sufficient variability in the data do exist, for example in cases in which comprehensive travel diaries are created through household surveys along a sufficiently long period. However, the SP surveys and the type of household survey above are very expensive means of data collection. Among the authorities who require travel demand models to evaluate policies and projects, the majority have no budget for obtaining high-quality data.

A high volume of demand modeling work is therefore done when the main source of information on existing travel comes from cheaper types of surveys, such as traffic counts. There is no systematic information about the options that individual travelers could choose from, the attributes of these options (cost, distance, level of service), or the characteristics of these individuals. With an effort, it is possible to convert the available data to a format suitable for modeling by using the standard methods mentioned above. But the result is a data set that lacks sufficient spatial variation and exhibits high correlation between attributes. As a result of the weaknesses of the data, the standard methods are often unable to identify a solution and the estimation process fails.

The failure of rigorous estimation techniques to use input data of a basic quality is therefore one incentive for proposing the approach described later. An additional incentive has to do with the limited ability to introduce, within the traditional methodologies, complex constraints on the sensitivity of the model. In England, authorities who seek funding from the central government for their transport interventions need to present demand forecasts created with models that comply with WebTAG (*3*). WebTAG is an Internet-based document that provides transport analysis guidance (hence TAG), including criteria that such demand models need to meet. WebTAG does not have a formal status outside England, but is seen as a useful source of advice, especially in countries that have not developed an equivalent set of modeling guidelines of their own. Such is the case in Ireland, where the approach presented here has recently been implemented.

The WebTAG criteria can be seen as part of a group of model validation criteria, which are presented in more detail later. The common feature of these criteria is the view that some preliminary knowledge about the way travelers behave is stronger than the evidence found in the data. The estimation work is required to create a model that exhibits plausible fit with the data, but only as long as some additional constraints are met, even if these may in fact contradict the data. The logic behind this is that these external constraints reflect knowledge that has been amassed over a large number of earlier studies and is therefore considered to be solid evidence on travel behavior.

Under the term "realism testing," WebTAG requires that the model be applied to calculate the demand elasticities to fuel cost, public transport fare, and travel time. A range of values for each of these elasticities is provided, and the model is considered unrealistic if the elasticities it implies are not within these ranges. It is recommended that if the model fails the realism tests, its parameters are to be changed and the tests are to be repeated.

WebTAG does not describe how the model should be estimated and does not deem any approach for estimating the model parameters inappropriate. It is explicitly stated that compliance with the realism tests should be given higher priority than accurate fit with observed demand or a robust estimation methodology. But finding a set of parameters that form a model complying with the WebTAG elasticities is not always easy. Even if the quality of the input data is sufficient to use traditional estimation techniques such as maximum likelihood, a common outcome is a model that implies elasticities outside the required range. The standard software for estimating logit models does not allow constraining the estimation procedure to accept only specifications that make a WebTAG-compliant model.

It is not the intention here to start a discussion on the logic in the range of elasticity values published in WebTAG. It is indeed debatable whether a preset range of elasticities based on past studies can be used to constrain the findings of a new study. However, there is no controversy about the formal (or informal) status of WebTAG in England and elsewhere. The need to satisfy the WebTAG criteria is therefore a political constraint at the heart of the model estimation process. Noncompliance leads to an undesirable conflict with the criteria for obtaining funding for transport projects.

The preset elasticity thresholds are not the only external constraint imposed on the demand model. In multilevel models, in which different demand responses form a hierarchy of submodels, WebTAG determines a range of acceptable values for the scaling parameters that define this hierarchy. As explained above, in many cases these values are based on rich experience and make good sense, but the point made here is that there is little room for the modeler's logical judgment even if they do not.

Other parties but WebTAG, in any country, add more constraints, which are generally driven by different opinions on what can be seen as logical travel behavior. A full list of these are presented later. Seeing these as constraints reflects the view that a model specification that does not meet them is infeasible, even if the statistical analysis shows otherwise. These constraints are commonly seen not as part of the model estimation problem but as informal criteria for testing whether the model makes good sense. The problem is that it is frequently found that the model does not meet one or more of these criteria, and it is then not clear what action should be taken. Attempts to improve the model such that it performs better with respect to one criterion are likely to make it perform poorly with respect to another.

In summary, the traditional model estimation techniques are often unable to solve the estimation problem if the quality of the data is compromised, and they also do not easily allow modelers to predefine types of travel behaviors that they or their clients believe cannot be true. Consequently, modelers in the United Kingdom and Ireland regularly build demand models by using a manual trial-and-error search. This is not only time-consuming, but is also unlikely to optimize whatever objective the modelers have specified. As mentioned earlier, the author is not aware of any widely used method for estimating travel demand models that accepts that the model should be influenced by such considerations. An attempt is made to formulate such a method here.

MULTIOBJECTIVE PROBLEM

The aim is to estimate a model by determining its parameter values. As input data is a set of origin–destination matrices (O-D). Some of the matrices contain information, by origin and destination, on the number of travelers choosing different travel options (e.g., different modes of transport), based on traffic counts, passenger counts, and similar sources. The remaining matrices contain information on attributes (i.e., variables) that are being considered for inclusion in the utility functions.

This is formulated as a multiobjective problem that features a market simulation tool and a solution search algorithm. The following paragraphs explain each of these terms and show that the approach is, essentially, pragmatic and very simple. It is also very flexible in the sense that it can easily be modified to include any set of objectives, any market simulation tool, and different solution search algorithms.

Stating that this is a multiobjective problem means that the choice of the best set of parameters is made by combining several objective functions, rather than one function as in a maximum likelihood process. One expression, which is called the meta-error, is not used to combine all objectives. The problem tries to minimize the value of this meta-error. All objectives are presented later.

Stating that the problem features a market simulation tool means that for each candidate set of parameter values, the full set of choices made by travelers from all O-D pairs is estimated. Then the different objective functions mentioned above are used to test to what extent these choice estimates meet the needs. That is done repeatedly with different possible parameter sets.

Stating that a solution search algorithm is used means that the aforementioned trial-and-error approach for testing multiple possible solutions of the estimation problem is replaced with a more systematic optimization technique. This helps in making an intelligent guess of what parameter set has a good chance of performing well on the basis of what is known on the performance of parameter sets that have already been examined. Many different techniques could be used for this purpose; the downhill simplex method, which will be elaborated on later, was chosen.

What is done practically is to undertake model estimation and validation simultaneously. Although the traditional practice is to reach a specification of the model and only then carry out validation tests, that may be of little use here. The model often fails some of the validation tests, and as a result of high correlation between the different model attributes, there are too many alternative combinations of parameter values equally likely to be the solution sought. By incorporating all the tests into one process from the outset, solutions that are not going to pass the validation tests later are overruled.

The following describes the different objectives used concurrently in this procedure:

- First set of objectives. Minimize the error in total trips for each alternative. The difference between the observed (i.e., from the input data) and the estimated (i.e., from the model) total number of travelers choosing each alternative is calculated as a proportion of the observed number of trips. The target value for this objective is of course zero. There are as many objective functions in this set as there are alternatives in the model.
- Second set of objectives. Minimize the error in the geographic dispersion of choices. Unlike the objectives from the first set, this set examines differences between observed and estimated figures for each O-D pair and then combines the errors across the entire study area, giving a higher weight to larger errors and to O-D pairs with a larger flow. To do this, the root mean squared weighted error (RMSWE) measure is calculated as follows:

$$\sqrt{\frac{1}{\sum y_i} \cdot \sum \left(\frac{x_i - y_i}{y_i} \right)^2 \cdot y_i}$$

where x_i represents the modeled number of travelers choosing a specific alternative for a specific O-D pair, i, and y_i represents the respective observed value for the same O-D pair. The ideal value of the RMSWE is zero, and again, the number of objectives in this set equals the number of alternatives in the model. Clearly, RMSWE could be replaced with many other measures or two-sample tests. Different tests have different sensitivities, which are widely described in the literature.

- Third set of objectives. Ensure that demand elasticities implied by the model are within the range recommended in WebTAG. This is, in fact, a constraint rather than an objective, but for convenience it is converted into an objective function by defining an expression (that the practitioner wishes to minimize) that includes a high penalty if the elasticities are outside the desired range. WebTAG specifies expected ranges for demand elasticity to fuel cost, public transport fares, and travel time. To estimate the model-based elasticity to each of these attributes in turn, the simulation tool calculates, for each O-D pair and for each candidate set of parameters, how the modeled choice would change if that attribute increased by 10%. The difference between this and the case without increase is summarized across the study area (weighted by the respective demand) to derive the average elasticity of demand to this attribute. Note that a 10% increase as a basis for the elasticity calculation is recommended by WebTAG. This recommendation is followed in studies with a large number of O-D pairs (typically, above 10,000). In studies in which the computational burden is less significant, checking how demand changes with different proportions of increase and decrease and then calculating an average elasticity across these is recommended.
- Fourth set of objectives. Ensure that scaling parameters are within the range recommended in WebTAG. A table in WebTAG sets out the expectations from the ratio of scaling parameters between each two adjacent components in the model hierarchy.
- Fifth set of objectives. Ensure that the proportion of each utility component (for each alternative) is within a logical range. The simulation tool contains a module that calculates, for each candidate parameter set, the contribution of each variable to the total utility of each alternative. For a specific O-D pair, the contribution of variable K (for example, walking time) is calculated as the parameter of K multiplied by the value of K, divided by the total utility. This is summarized across all origins and destinations, and the demand in each pair is used as a weight. The idea is to ensure that the automated model estimation process does not let the relativities between the utility components contradict one's intuitive judgment. Thus, the objective here is to minimize an expression that includes a high penalty on peculiarities such as when the walking time component constitutes 80% of the total public transport utility, or when travel time constitutes 10% of total car utility.
- Sixth set of objectives. Ensure that values of time derived from the model are within a logical range. Practitioners are usually sufficiently familiar with their study areas to speculate within what range the values of time derived from the model must lie. That range in the solution search algorithm is used to reject model specifications that contradict practitioners' local knowledge.
- Seventh set of objectives. Ensure that alternative-specific constants derived from the model are within a logical range. While working on other models practitioners develop an understanding of the logical range of alternative-specific constants. Again, this knowledge is used formally in the algorithm to ensure that this automated procedure does not accept illogical solutions.

As explained earlier, what practitioners try practically to minimize is a meta-error, that is, a weighted average of the values of all objectives. To calculate the meta-error a weight is assigned to each objective; these weights have two roles. First, they convert the different objectives, which have different units or scales, to a uniform (albeit abstract) scale. Namely, the weights correct the imbalance caused by the fact that an improvement of 0.1 in one objective is not equally significant to a similar improvement in another objective. Second, the weights also reflect the view about which of the objectives are more important to optimize. For example, the weights can be used to make the process more sensitive to the demand elasticities and less sensitive to the goodness of fit with observed demand.

Setting up the weights at logical levels ensures that the meta-error is more affected by the objectives that have not yet reached satisfactory values. To explore the effect of each objective on the meta-error, charts such as the one illustrated in Figure 1 are used. The chart displays the composition of the meta-error; it is affected by the magnitude of the various objectives and by their weights. It is suggested that before the estimation process is run the weights be determined by using any initial specification of the model, so that the relative sizes of the weighted objectives match the practitioner's interpretation of the size of the problem that each one of them indicates. The initial model specification used for this purpose can even be one in which the parameters have been chosen randomly. It is recommended that this exercise be repeated several times before the weights can be deemed suitable for the estimation process.

Charts such as the one in Figure 1 are used not only at the beginning of the estimation process, when specifying the weights, but also throughout the process, to monitor what features of each candidate

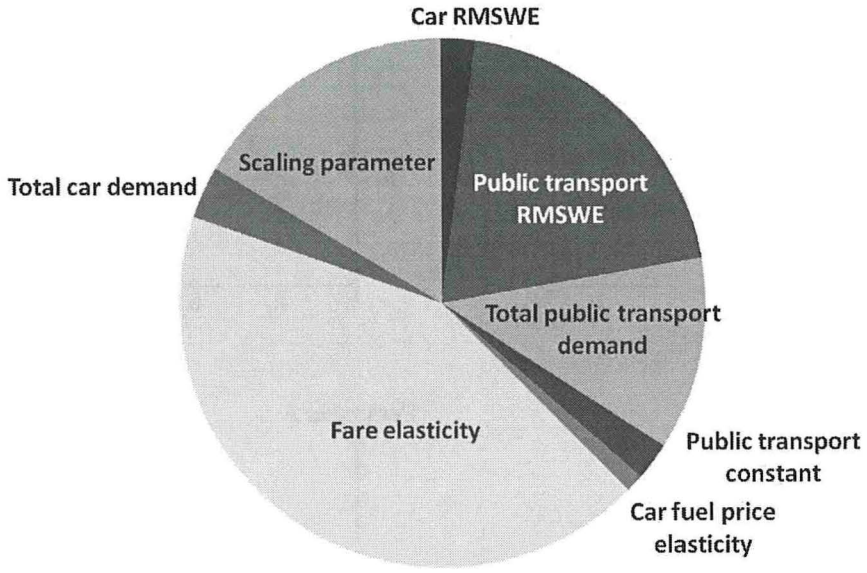


FIGURE 1 Composition of the meta-error.

model specification are those that most affect its performance. The chart does not look at the absolute size of the meta-error; that is dealt with separately by the algorithm presented in the next section.

When reviewing the composition of the meta-error, the modeler may decide at any stage that a change of weights is required to place more emphasis on specific objectives. There is also a case in which the process changes a weight automatically. Such a step is introduced if one of the objectives for which the target value is merely a guess reaches its optimal value. This might be an indication of overconstraining, and the procedure therefore reduces the weight for the respective objective. The objectives for which the optimal value is merely a guess include elasticities, scaling parameters, and constants, as opposed to the error in total demand, for which it is known with certainty that the target is zero.

DOWNHILL SIMPLEX METHOD

The search for a solution for the model estimation problem uses the downhill simplex method, which has been previously used for a range of calibration problems (*4*). It is not a particularly efficient optimization technique in regard to the number of iterations required, and it also does not guarantee convergence to the global optimum. A further disadvantage is that it does not generate standard statistical measures such as a *t*-test for different attributes.

Nevertheless, the downhill simplex method has several merits that make it suitable for the present needs:

- It is suitable for optimizing objective functions that do not have a closed form. The present parameters are used in a logit model in many O-D pairs, and the output demand is aggregated across all pairs to calculate the objective values; hence, the meta-error is not an explicit function of the parameters.
- The method does not use derivatives, which would take many hours to calculate numerically in this case.
- The quality of the data will clearly affect the quality of the model, but failure to estimate any model (which is a possible outcome when maximum likelihood is used) is unlikely.
- The method is easy to program. Versions of it have been created by using different software packages, including C++, Visual Basic for Excel, and Stata. The Excel version is the most transparent but runs slower. The Stata version lacks this transparency but is much faster.

The following is a basic description of key concepts of this algorithm. A simplex is a geometrical shape in a multidimensional space. At each corner of a simplex there is a vertex; in an N-dimensional simplex there are N+1 vertices. When the downhill simplex method is used to optimize the values of N parameters, an N-dimensional simplex with N+1 vertices is used.

Each dimension of the problem represents one parameter, and each vertex is one candidate set of values of all the parameters. The simplex at each stage of the process is the best group of N+1 candidate solutions the practitioner is aware of at that point.

A simple problem, with two parameters to estimate, is illustrated in Figure 2. Because there are two parameters, the simplex is simply a triangle. Each of the three vertices of the simplex is a possible solution of the problem. The coordinates of Vertex 1, A_1 and B_1, are the respective values of Parameters A and B according to Solution 1; the same goes for Vertices 2 and 3.

The fact that each vertex is a possible solution of the model estimation problem means that each vertex is in fact a model, for which the practitioner can calculate the difference objectives and the meta-error. At the first stage of the process the objective values for all vertices must be calculated. After the first stage, an iterative loop is started, and in most iterations the number of times one needs to calculate the objective value per iteration is much smaller.

The core iterative process works as follows. Because the objective value (i.e., the meta-error) for each vertex is known, the vertex that has the worst (i.e., highest) value can be identified. This is deemed the worst vertex, and the other vertices (the remaining two, in the triangle example) are deemed the base of the vertex. To eliminate the worst vertex and replace it with a better solution of the problem, a reflection maneuver is undertaken. Namely, the worst vertex is replaced with a point that lies at the same distance from the

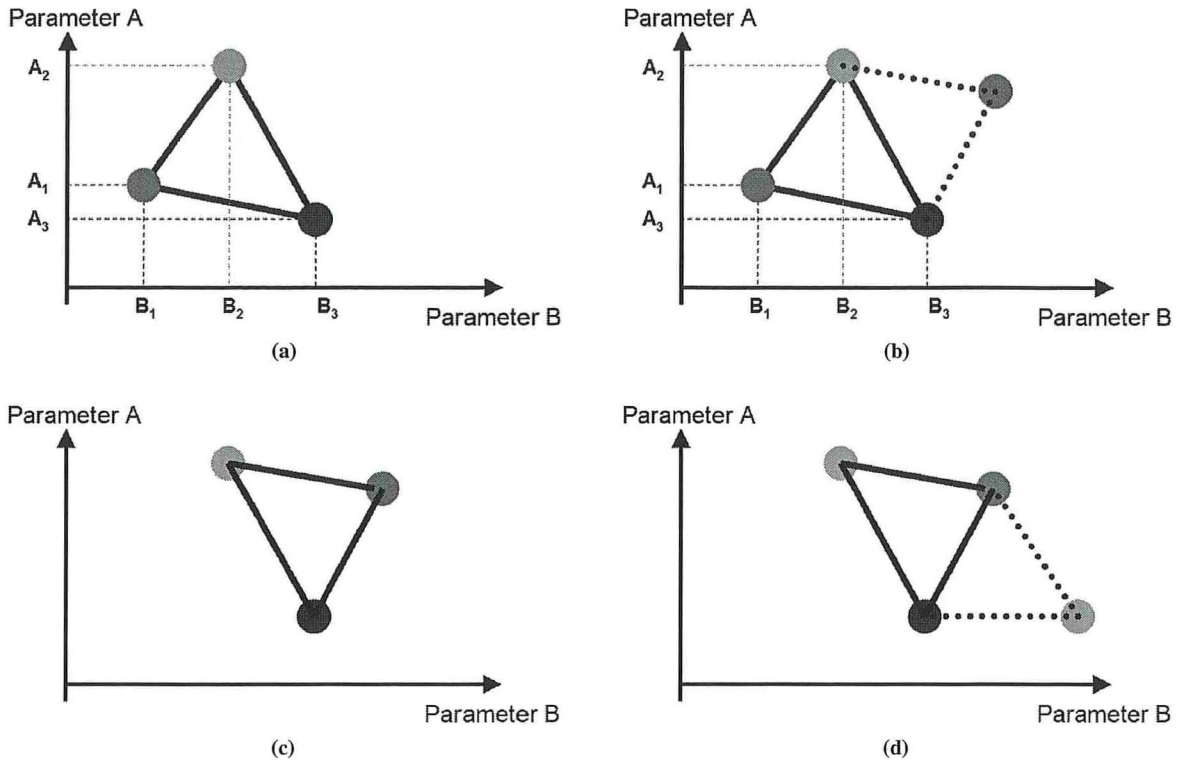


FIGURE 2 Downhill simplex method: (*a*) Stage 1, (*b*) Stage 2, (*c*) Stage 3, and (*d*) Stage 4.

base but on the opposite side. In Figure 2, Stage 1 identifies Vertex 1 as the worst vertex, and Stage 2 demonstrates a reflection maneuver. This is merely a guess that the new vertex, with the implied new values of the parameters, would have a better value of the objective, but it is an intelligent guess that often works well. If the objective in the new location is indeed better than the worst objective value in the original simplex, this maneuver is completed and one now has a new simplex, presented as Stage 3 in Figure 2. Stage 4 starts a new iteration of a similar nature.

The reflection maneuver does not work well in every iteration. If a simple reflection does not lead to an improved objective value, some alternative types of reflections are investigated. A full description of all technical aspects of this process is avoided here, but they are available in the optimization literature (*5*).

DUBLIN CASE STUDY

The approach described here was recently applied in a number of studies in the United Kingdom and Ireland. The case of the upgrade of the Dublin Transportation Office (DTO) model, completed in May 2009, is presented here.

DTO is an agency responsible for the strategic planning of the transport system throughout the Greater Dublin Area. Its comprehensive model required upgrading following many recent demographic, geographic, and economic changes in and around Dublin. The demand modeling work used much data collected between 2006 and 2009. Although the data did include rich sources of information such as household surveys, they still exhibit the typical problems of RP data (as discussed earlier). This meant that the most reliable information on travel demand was a set of matrices that had first gone through an extensive series of corrections using traffic and passenger counts.

The DTO model has generation, distribution, mode choice, time-of-travel choice, and route choice components. The task presented here focused on the mode and time choice components. Each of these two submodels has its own internal nested structure, in addition to the nesting implied by the overall model hierarchy. Estimates of journey times, costs, and other attributes are based on a highway assignment model (using the Saturn software) and a public transport assignment model (using Trips). Figure 3 depicts the model structure.

The time choice model deals with the choice between the 3 h of the morning peak period. In the assumed model hierarchy (which is later confirmed by the results), time choice sits just above route choice; hence the time choice model has the most detailed utility functions. For car users, these functions include the traveled distance, travel time, and any tolls if applicable. For public transport users, the utility functions include the in-vehicle time, waiting time, walking time, number of transfers, fare, and the amount of time spent in crowded conditions. Time choice is not modeled for slow modes (i.e., walking and cycling).

The mode choice model is located above the time choice, and therefore the utility functions for car and public transport include the composite utilities across the different times of travel and a public transport constant. For slow modes, the utility function includes the time and a constant. The estimation of both models also includes determining the value of structural scaling parameters.

The process distinguishes between travelers with and without a car available for the journey. Models are estimated separately for five different home-based journey purposes (commuting, business, education, shopping, other) and for non-home-based trips. To build all these models, the algorithm described earlier was set up as a Stata code. Because

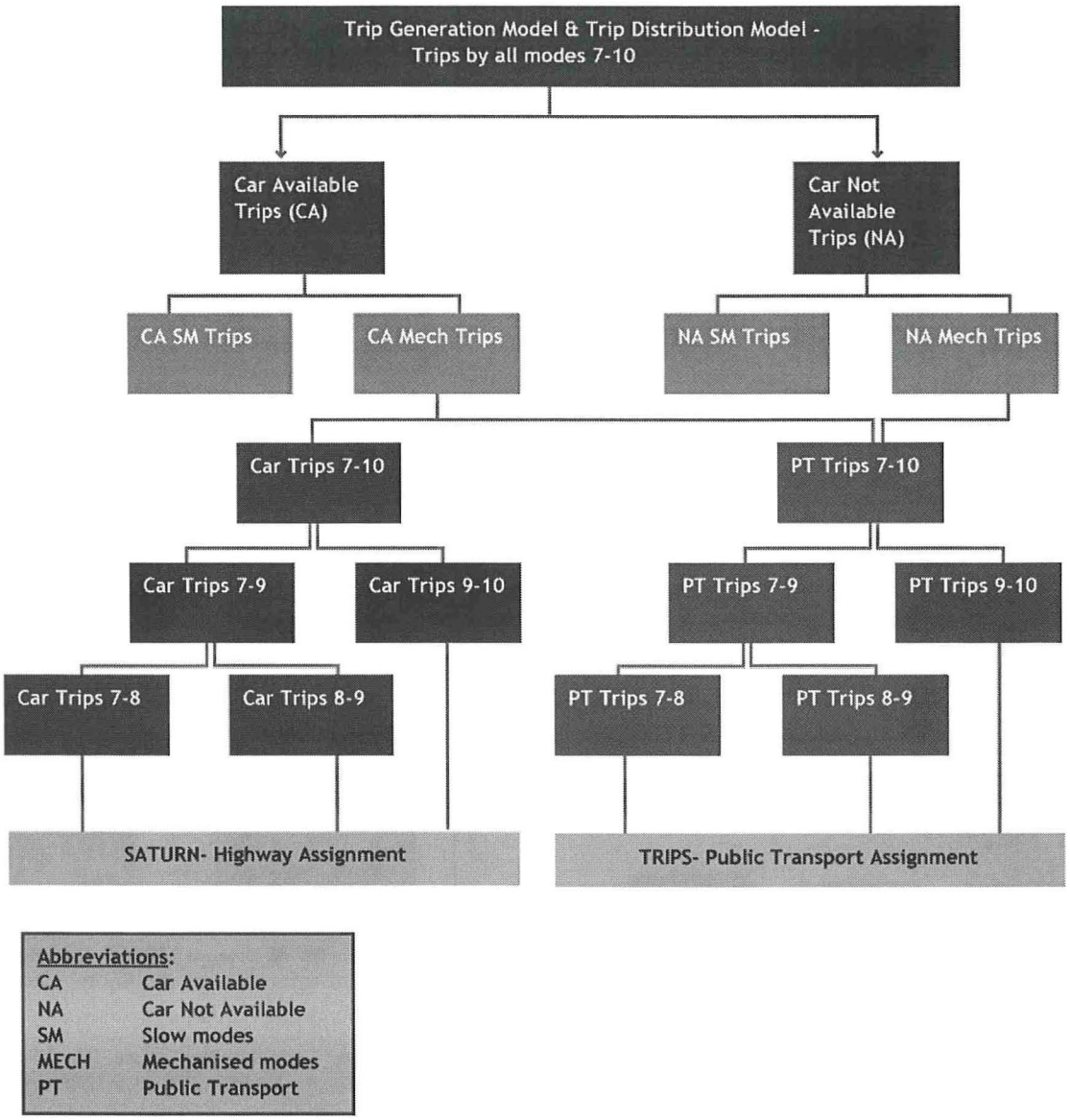


FIGURE 3 Structure of DTO model.

of a lack of space, a full technical configuration of the algorithm is not presented here. The algorithm includes the following:

- Weights for the different objectives,
- Recommended ranges of elasticities from WebTAG,
- Recommended ranges of scaling parameters from WebTAG,
- Threshold values of time for each journey purpose, and
- Likely ranges of alternative-specific constants.

Because there are models for different journey purposes, travelers with or without a car available, and a large number of objectives, the full set of results cannot be presented within the limited space here. Full outputs are available on request from DTO or the author. The performance of the models estimated for one journey purpose are illustrated here, namely, home-based commuting, and mainly the first three sets of objectives described earlier are discussed. It is stressed, however, that the extent to which all objectives were met is far more satisfactory than what is normally experienced when the estimation process does not formally account for these multiple objectives.

Figure 4 shows how the time choice models and the mode choice models for commuters perform in estimating the total demand per alternative, as defined in the first set of objectives. In each part of this figure, the observed number of travelers (based on the input data) is compared with the model output. It is easy to observe that the model is successful in approximating the overall attractiveness of the alternative times of travel and modes.

Figure 5 shows the results with respect to the second set of objectives, that is, the accuracy of the demand estimates at an individual O-D level. The full analysis includes a series of values of the RMSWE, but to keep the illustration here intuitive, the same information is displayed graphically.

In each of the scatter plots in this figure, each point refers to an individual O-D pair. The point is located so that the observed demand

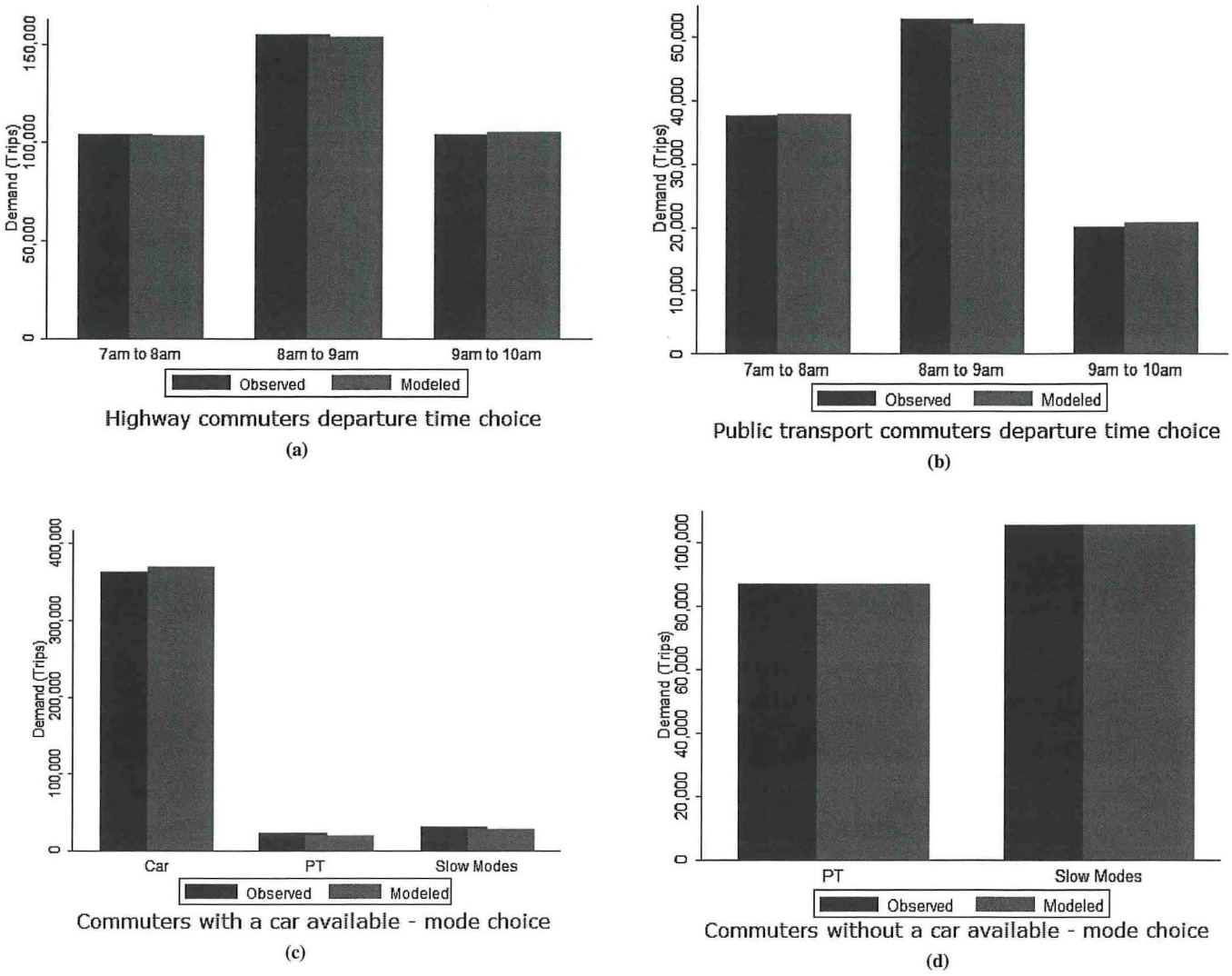


FIGURE 4 Results for commuters: total demand (Objective 1) by choice—(*a*) highway commuters departure time choice, (*b*) public transport commuters departure time choice, (*c*) commuters with a car available—mode choice, and (*d*) commuters without a car available—mode choice.

determines its horizontal position and the modeled demand determines its vertical position. The extent to which these points lie on a 45-degree line from the origin indicates how well the model fits the observed data. There will always be a "cloud" of points, because no model will perfectly predict reality at such a high level of geographic detail, but a relatively narrow cloud that symmetrically straddles the 45-degree line will indicate a good level of fit with no systematic bias.

In general, the scatter plots show a very good degree of association between the observed and modeled values. The accuracy of the model is reduced, though, in instances in which demand itself is very low. The modal shares of public transport and of the slow modes form a very low proportion of the total demand, and the ability to reproduce these is compromised. Note, however, that modeled–observed demand comparison at the individual O-D level is a rigorous test that is normally not required by any formal standard. Experience shows that with whatever model estimation technique, it is always difficult to obtain a high level of model fit for the low-incidence alternatives. These results are therefore deemed satisfactory in the current context.

Figure 6 presents the elasticities of home-based commuting demand, calculated from the model. WebTAG requires that fuel price elasticities be in the region of −0.1 to −0.4, depending on journey purpose; for commuters it is expected that elasticities are less negative than for more discretionary types of trips. The model is very successful in confirming that the expected elasticities apply to the population of Dublin. The fare elasticities from the public transport commuters' model are very low, but that appears reasonable given that a large proportion of public transport commuting trips are made by travelers with no access to a car. The travel time elasticities from the model are well within the WebTAG range, too.

Although these cannot be presented here in full, it was found that scaling parameters, values of time, alternative-specific constants, and the proportion of the utility formed by each attribute all met the preset criteria. Similar results were obtained for other journey purposes.

Further validation of the plausibility of the parameter estimates was achieved by plotting a series of maps describing the composite utility values across all travel options, calculated with the set of parameters suggested by the multiobjective process. This allows practitioners to examine whether the relative ease of access between

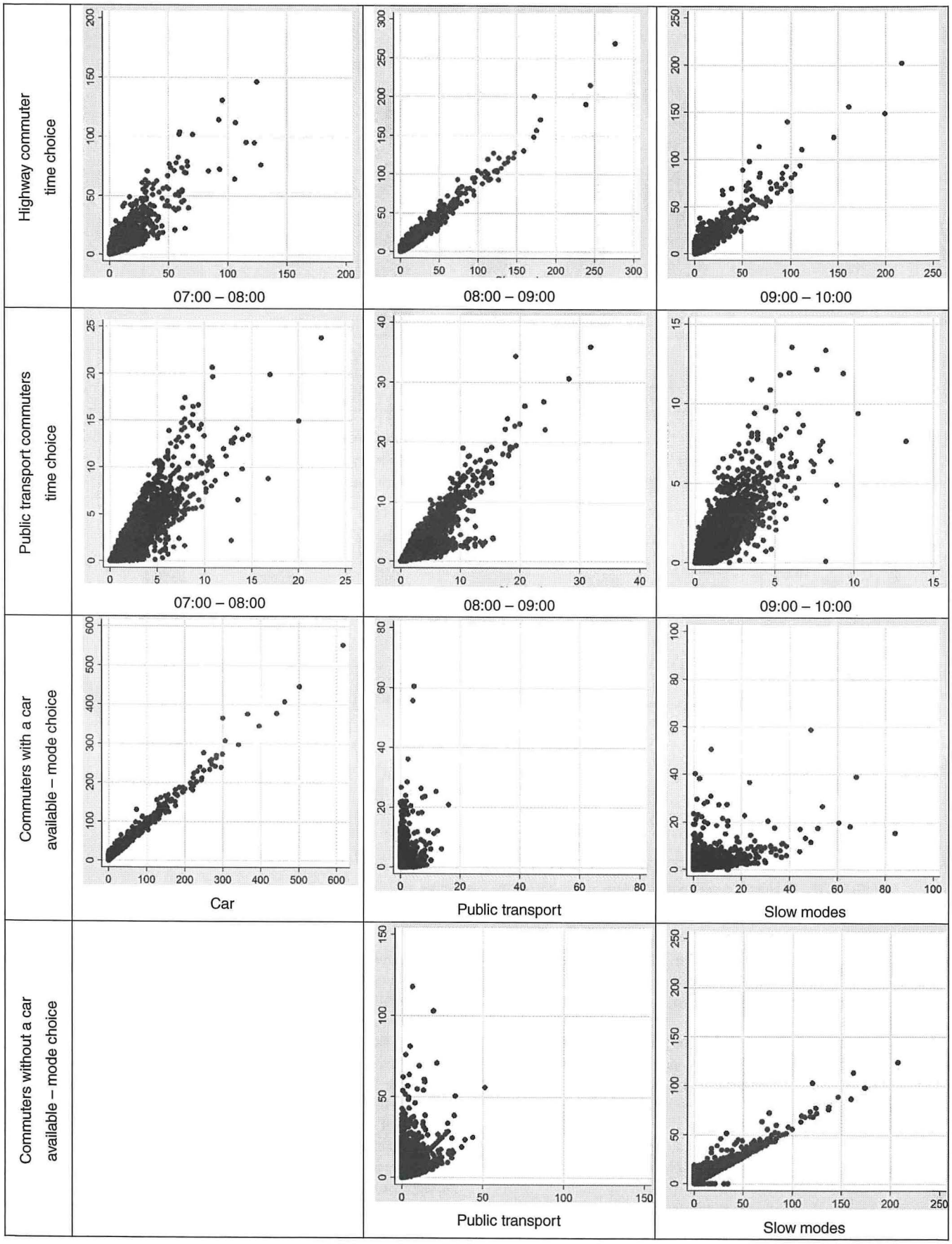


FIGURE 5 Results for commuters: demand by origin and destination (Objective 2).

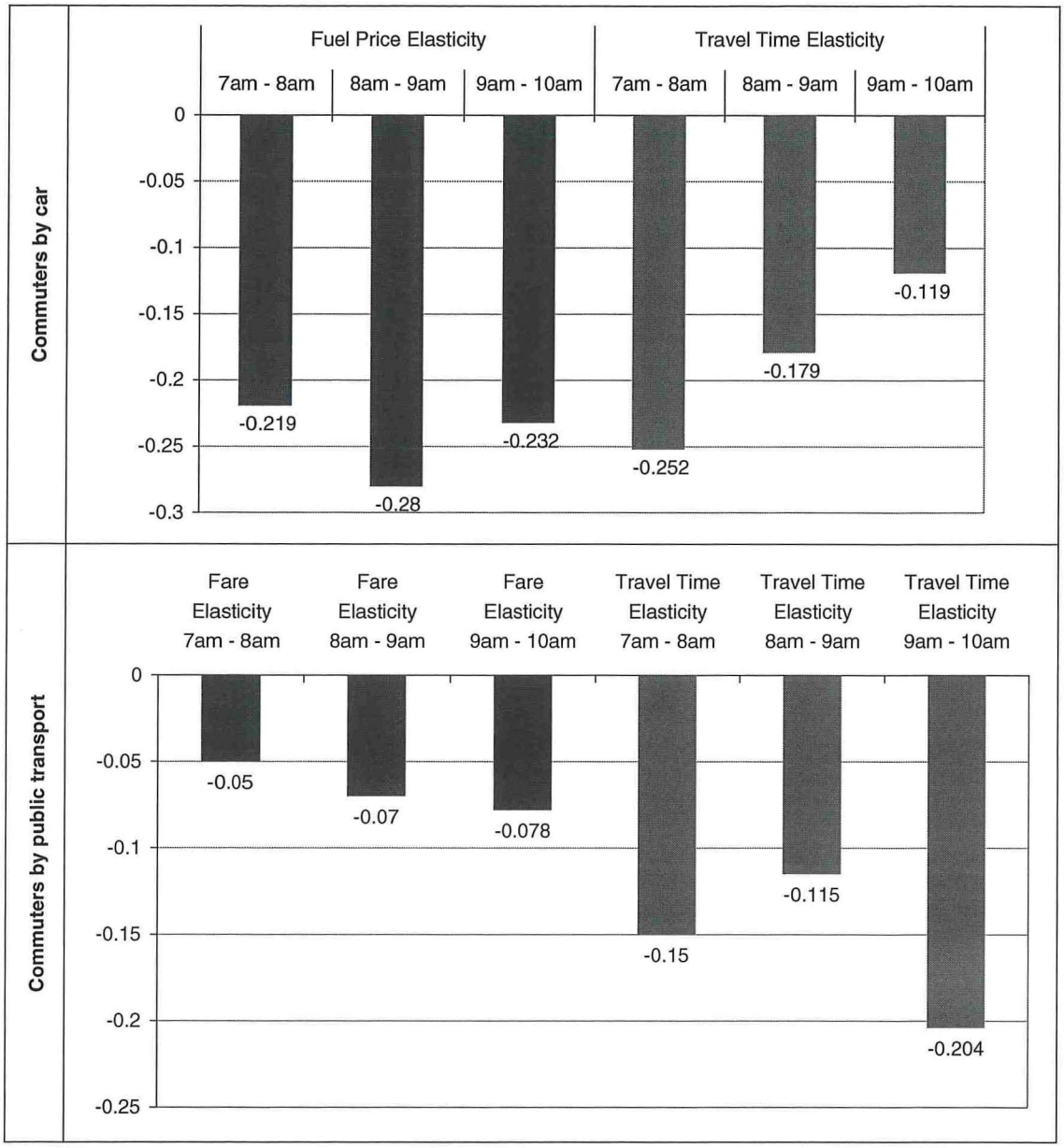


FIGURE 6 Results for commuters: elasticities implied by the model (Objective 3).

different areas, according to the model, matches their informal knowledge of the Greater Dublin Area. Figure 7 demonstrates this type of additional validation with a map of the utility values from all origins to the Dublin Airport.

It was generally found that the multiobjective process helped in identifying a model that satisfied the standards of statistical and behavioral analysis and, at the same time, complied with constraints determined by external parties.

CONCLUSION

A pragmatic approach to the estimation of travel demand models was presented; it differs from the traditional techniques in that it bases the solution on a number of objectives. These objectives include classical measures of model fit, but also other tests that reflect what the modelers or other stakeholders believe can count as a credible model.

The Dublin case study presented here confirms that this procedure can help identify a model that jointly meets many of the expectations that different parties have of it. This is by no means meant to criticize the traditional estimation approach, which is indeed based on stronger theoretical foundations. But it illustrates that a pragmatic approach for those cases in which the traditional approach fails (because of either data quality or political constraints) is available and works efficiently.

In future publications the intent is to present further work based on the approach presented here. That includes comparison of models estimated with this approach with equivalent models created by using the traditional practice. It also includes assessment of models estimated with the proposed approach with different sets of objectives

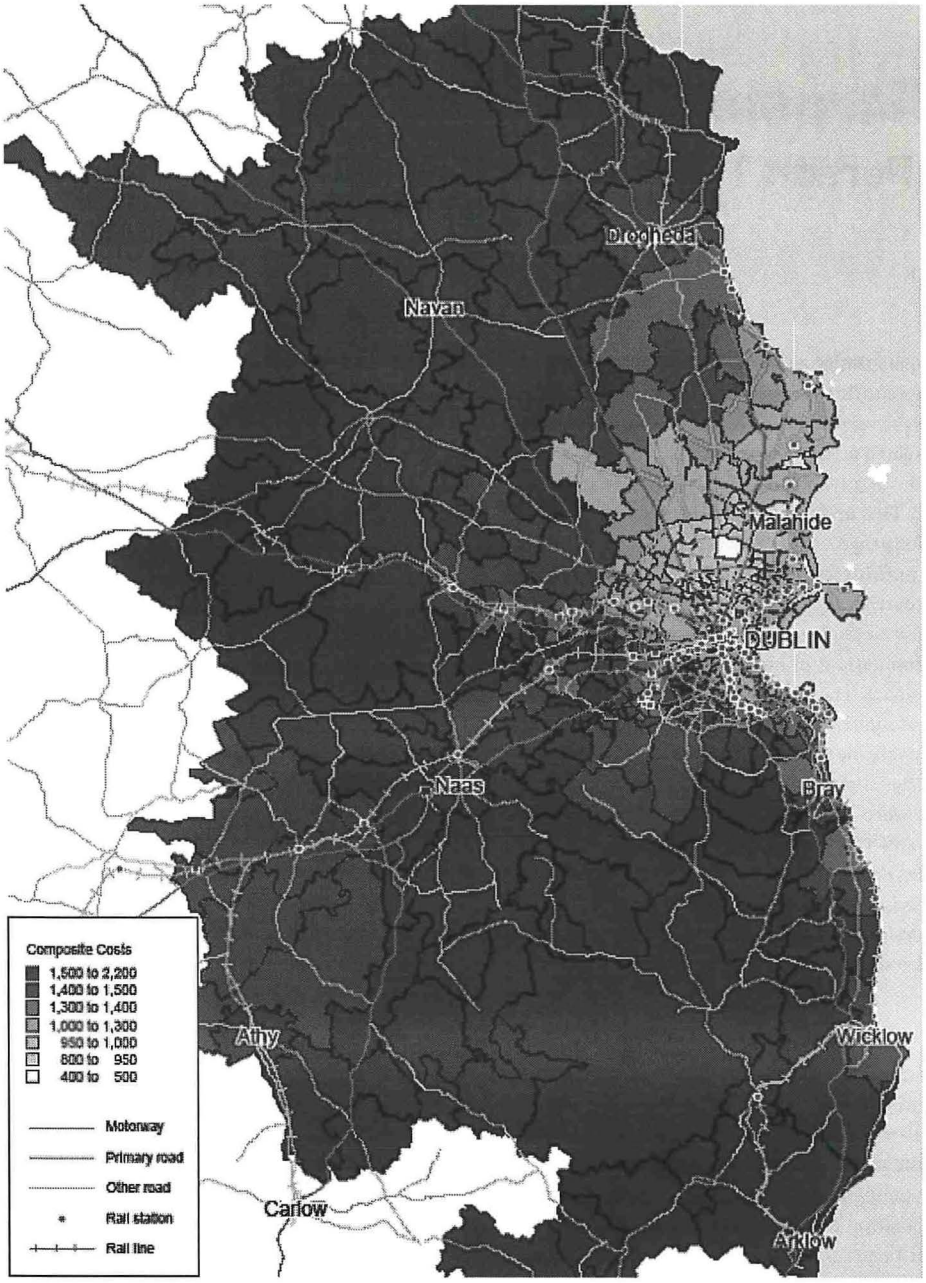


FIGURE 7 Utility of travel from all origins to Dublin Airport.

and weights, and some consideration of the risk of bias. Contribution from others to this discussion would be valuable, too. A suggested area for such work would be an attempt to incorporate a maximum likelihood objective as part of the multiobjective tool so that the solution could benefit from the advantages of both approaches.

REFERENCES

1. Ben-Akiva, M., and S. T. Lerman. *Discrete Choice Analysis: Theory and Application to Travel Demand.* MIT Press, Cambridge, Mass., 1985.
2. Train, K. *A Comparison of Hierarchical Bayes and Maximum Simulated Likelihood for Mixed Logit.* 2001. http://elsa.berkeley.edu/~train/compare.pdf. Accessed June 5, 2006.
3. *WebTAG, Transport Analysis Guidance.* Department for Transport. Dublin, United Kingdom. http://www.dft.gov.uk/webtag/.
4. Hollander, Y., and R. Liu. Estimation of the Distribution of Travel Times by Repeated Simulation. *Transportation Research Part C,* Vol. 16, No. 2, 2008, pp. 212–231.
5. Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C.* Cambridge University Press, New York, 1992.

The Transportation Demand Forecasting Committee peer-reviewed this paper.

From Microsimulation to Nanosimulation

Visualizing Person Trips over Multiple Modes of Transport

Gordon Duncan

In the past 15 years microsimulation software tools have increased the ability to analyze congested traffic by modeling at the level of individual vehicles. Although this can be very detailed, it assumes that mode choice is fixed. The research presented models the people in the network, either walking or in vehicles, following each person for an entire trip through multiple modes of travel. This approach is called "nanosimulation." The paper presents a pilot project that analyzes access to an airport by comparing multiple parking options, rail transit, drop-off, and taxi access. Generalized cost incorporating time, distance, and price is visualized for each access method and allows comparison of the total end-to-end cost of all combinations of modes in an interactive three-dimensional simulation model. The primary objective of the research presented is to prove that analysis at this level is practical and can provide insight that is not available from other methods. A historical perspective of microsimulation is presented to illustrate how seemingly impractical models that were run on a supercomputer in 1994 are now in common use on sub-$1,000 everyday computers. The paper describes the technologies, data structure, and algorithms used in the research and addresses the issues of data availability and model repeatability. The model area is described, with illustrations of the visualization. The final sections present results of the project, followed by conclusions and recommendations.

Macrosimulation (or macroscopic simulation) software tools model traffic on a transportation network as a time-varying value of flow for each link or lane and are often used for building strategic, wide-area models. Microsimulation (or microscopic simulation) tools aim to simulate traffic by modeling hundreds or thousands of individual vehicles, changing the position of each one at a fixed time interval within a predetermined time period. The vehicles are moved according to a set of rules governing motion, interaction, and constraints and controls applied by the network. For example, a model might simulate 25,000 vehicle trips during a 4-h morning peak by using a time-step interval of 0.5 s, modeling freeways, surface streets, traffic signals, and so forth. Many such models simulate public transit vehicles in addition to private vehicles, and some have recently added the ability to model pedestrians and their interaction with vehicles. These software tools allow the analyst to assess the performance of transportation networks, both existing and proposed. It is important to recognize that each software tool has its own capabilities and limitations, and none is perfect. It is the responsibility of the analyst to ensure that a suitable software tool is selected for any given task of assessing transportation network performance and that the tool is then calibrated to local conditions. Although a software package may provide default settings for its parameters, it is not standard practice to use these; state and national transportation authorities frequently produce guidelines on such calibration (*1–4*).

Although a microsimulation model can be very useful, it tends to focus on vehicles and the traffic load that they create, rather than the people in those vehicles and the underlying travel needs of the people. This paper focuses on research that aims to analyze network performance at this more detailed level: that of the end-to-end trips made by people over multiple modes, rather than single-mode trips made in a vehicle or on foot. The term "nanosimulation" (a contraction of "nanoscopic simulation") has been proposed for analysis at the detailed level of the individual person, including the dynamic decision processes in that person's mind (*5, 6*). This paper further defines nanosimulation as a process that incorporates the dynamic mode-choice function of a person into a detailed model, so that an individual person in the model may make instantaneous choices between available modes as well as choices between available routes. For example, a person released into the model may initially intend to walk to a bus stop and take the bus across the city center. On arriving at the bus stop, the information display on the bus stop may advise that the bus is running 10 min late as a result of traffic congestion in the model, and the person will then decide to hail a taxi or walk to a taxi rank and take a taxi to the destination. Although it is fairly common for a microsimulation tool to model dynamic route choice within a mode, its demand is commonly specified by an (O-D) matrix of mode-specific trips, making it impossible to model a person dynamically switching from one mode of transport to another. A nanosimulation model can represent dynamic mode switching by allowing each individual agent to choose a new mode of transport during its trip in instances in which this choice has been triggered by congestion or other delays causing a change to the total cost of travel using a generalized cost equation. Current commercially available microsimulation models, such as Paramics, VISSIM, and Aimsun, do not allow mode switching; demand by mode is a fixed input parameter. Although some existing packages, such as TRANSIMS, model trip generation at the individual person model, the simulation component models the supply of transport at the vehicle level only and is thus classified as a microsimulation.

A nanosimulation tool such as the one described here is best suited to the modeling of small- to medium-size networks. Typical usage scenarios might be the assessment of access to airports, park-and-ride facilities, and public transport interchanges or provision of new cycle lanes.

Azalient, Stirling University Innovation Park, Stirling FK9 4NF, United Kingdom. gordon.duncan@azalient.com.

Transportation Research Record: Journal of the Transportation Research Board, No. 2175, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 130–137.
DOI: 10.3141/2175-15

TABLE 1 Performance and Cost of Microsimulation Over Time

Year	Hardware	Cost (US$)	2009 Cost (US$)	N	S	t	P	Q
1994	Cray T3D	12,000,000	17,465,000	125,000	1.0	0.5	125.0	139,720.0
1996	SGI Indy Workstation	10,000	13,747	5,000	1.0	0.5	5.0	2,749.4
1999	Desktop PC	3,000	3,969	5,000	2.4	0.5	12.0	330.8
2004	Desktop PC	1,500	1,712	5,000	7.2	0.5	36.0	47.6
2009	Laptop PC	750	750	5,000	24.0	0.5	120.0	6.3

NOTE: PC = personal computer.

HISTORICAL BACKGROUND ON SPEED AND COST-EFFECTIVENESS OF SIMULATION

When a simulation as detailed as the nanosimulation model described is proposed, the practicability must be addressed: How much time is required to run a useful model, and what is the cost of suitable computing hardware?

Early microsimulation software tools required vast resources. In 1994, an early version of one, Paramics, ran a large model on a Cray T3D supercomputer in the University of Edinburgh (*7*). This room-size machine was valued at approximately US$12 million. Skeptics cast doubt on microsimulation ever being a practicable solution for transportation analysis. Less than 2 years later, it was possible to run a neighborhood-size Paramics model on a US$10,000 graphics workstation at a speed equal to real time (*8*). Currently, a similar model can run many times faster than real time on a laptop computer that costs US$750, and microsimulation has been adopted by many, if not most, transportation authorities. This section aims to illustrate the time required for a simulation modeling technology to move from the research arena to the practical use arena.

When models are compared across a range of hardware and software, it is useful to define a performance index to normalize the most significant parameters. The most significant parameters determining the speed of the simulation are the number of vehicles (or agents) in the model and the time interval. A useful measure of performance is the "speed-up," the ratio of elapsed simulation time to elapsed real time.

The performance index, P, is defined as

$$P = \frac{N \times S}{k \times t}$$

where

$$S = \frac{T_S}{T_R}$$

N = number of vehicles in the model,
S = speed-up,
t = simulation interval,
T_s = elapsed simulation time,
T_R = elapsed real time, and
k = 2,000, a normalizing scale factor.

The cost index, Q, is defined as U.S. dollars per unit of performance:

$$Q = \frac{C}{P}$$

where C is the cost of hardware, adjusted to current day using the U.K. Consumer Price Index.

Table 1 and Figure 1 show that the cost per unit of performance has fallen significantly during the past 15 years, and it would be reasonable to expect that it would go on falling. By implication, the performance of a simulation tool for a fixed cost of hardware

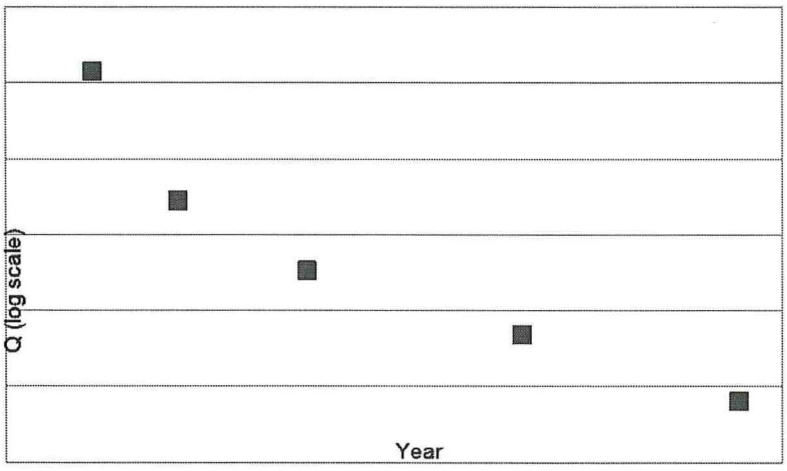


FIGURE 1 Change in cost index of microsimulation over time.

has risen steadily. This increase in performance is very similar to the performance improvement suggested by Moore's Law, that of doubling every 2 years.

OBJECTIVE

The primary objective was to identify whether it was feasible to create a nanosimulation model and run it successfully, within a reasonable time to produce results suitable for analysis. Here, a nanosimulation model is one that is capable of modeling all the people in a transportation network from their origins to their destinations through any number of mode changes. Secondary objectives were to build a tool that emphasized ease of use, to speed up the process of building a model, and to build a tool that could be extended by its users, by providing an application programming interface (API).

The software tool created was named "Commuter" because in many cases the people it modeled would be commuters, and also, it aimed to shorten (commute) the process of building a model. To address the issue of ease of use, Commuter employs advanced visualization technologies to represent the three-dimensional (3-D) model space and provides a user interface designed to allow fast navigation around this space. These technologies and the provision of an API are discussed below.

SOFTWARE TECHNOLOGIES

Many advances in computing technology originate from demand for performance from the field of computer gaming. Modeling of transportation networks can benefit from these advances in software and hardware. In software, Commuter uses Java and Open Graphics Library (OpenGL) development platforms that allow rapid creation of user interfaces and 3-D graphics, respectively. Both of these development platforms count computer gaming as one of their principal application fields. In hardware, Commuter can take input from a joystick or a multiaxis mouse to allow the user to navigate more easily around a 3-D model (*9*).

The Java software platform is an open-source product distributed by Sun Microsystems. It is more than just a programming language—it is a set of tools that allow rapid development of software applications. Java is designed to ensure that an application is hardware neutral. The phrase "write once, run anywhere" is used to summarize this concept, meaning that the application can be transferred from one type of device to another, with very little modification. So an application that is written in "pure" Java on a Microsoft Windows personal computer (PC) could run on a mobile phone or a supercomputer with little or no modification. In practice, Java applications often use small modules written specifically for a particular platform to improve performance in critical areas, and these so-called "native" modules must be translated for the application to work on another device. However, the fact that the vast majority of the code can be used without change remains a persuasive argument for Java.

Java's performance has improved substantially since the early versions. Despite a lingering myth that it is slow, performance of Just-In-Time compilers relative to native compilers has, in some tests, been shown to be quite similar, or even faster (*10*).

OpenGL is a standard specification defining a cross-language, cross-platform API for writing applications that produce 2-D and 3-D computer graphics. The interface consists of more than 200 different function calls that can be used to draw complex 3-D scenes from simple primitives. OpenGL is not part of the standard Java platform, but several wrapper libraries exist to allow calls to OpenGL to be made from within Java. Two of these were assessed for use in this project, the "official" solution of Java bindings for OpenGL (JOGL) and the alternative Lightweight Java Game Library (LWJGL). It was found that the performance of LWJGL was slightly better, and that is used as the default.

USER EXTENSIONS THROUGH AN API

No simulation modeling software tool can ever contain all the functionality that might be required by every user. The provision of an application programming interface or API allows any user with programming skills to extend the application according to the needs of the project. Commuter provides a Java-based API, documented in Javadoc, the standard format produced by Java's built-in API documentation generator.

The Commuter API has a hierarchical structure that mirrors the hierarchy of the data file describing a network. At the top level, the user can load a model by its filename and request access to a handle for that model. This top-level handle can then be queried for a range of second-level handles, such as Network, Demand, Assignment, or Results. Each of these second-level handles can then be queried for progressively more detailed information, down to the level of individual objects in the model. Using Network as an example, the user can request the set of all intersections in the network, select an intersection by name, and request a list of all pedestrian crossings on that intersection and the signal group to which each belongs. The user's own algorithm could then set parameters for when those pedestrian crossings are activated, changing the performance of the network.

CHALLENGE OF REPEATABILITY

A nanosimulation model requires many input parameters, even more than for microsimulation. A specific challenge to be addressed is how to build the software tool in such a way as to make model results repeatable for fixed input parameters. A related, and more difficult, challenge is to build the tool such that a small change in one area of the model does not cause large changes in the results for trips through the model that are not directly affected by the small change to the network. For example, if the model represents part of a city center and a freeway that bypasses the city center, a change to the parking arrangements in the city center should not cause large changes to the trips that use the freeway only. Although this may sound obvious, experienced users of microsimulation software will recognize this as one of the pitfalls of a system that uses quasi-random number generators. It is commonly the case that a change to the network can cause a change to the order and type of trips generated, thus greatly changing the outcome of the model, a variation of the so-called butterfly effect. Although it can be argued that this is valid for any model that uses a random element to initialize its variables and that multiple runs of the model should be aggregated to smooth

out variations, it is nevertheless very useful on a practical level to minimize this type of variation if at all possible.

That challenge was addressed by decoupling the demand specification from the trip generation. In many microsimulation models, trips are generated at simulation time by using a Monte Carlo method. That is, in every time interval of the simulation, a number from a pseudorandom number generator (PRNG) is used to select a value from a distribution, and if this value exceeds a threshold defined by the demand specification, then a trip is generated. However, if the network is modified, even only slightly, the sequence of numbers drawn from the PRNG may change, causing a different sequence of trips. This variation in the sequence of trips can lead to large changes in the performance of the network. For example, a single vehicle turning across oncoming traffic at a busy intersection can bring about a marked change in the congestion, causing large variations to throughput in a particular run of a model. In Commuter, a table of randomly generated trips is built once and then stored with the network definition. The same table of trips can be used on the base network and on the modified design network, making it easier to compare the results. Although storing the trip table with the network requires the data file describing the model to be larger, data compression is used to minimize the size of the file.

BUILDING A NANOSIMULATION MODEL

Objects Required to Represent the Network

Many of the objects required for the network are similar to those in a microsimulation network, but in some cases defined by more detailed parameters. Network objects define the surfaces on which the agents move. These surfaces include lanes and intersections for road vehicles, track for rail and trams, and walking surfaces for pedestrian movement. Pedestrian crossing objects are required where these surfaces intersect, and signalized and unsignalized crossings should be supported.

Objects Required to Allow Change of Mode

It is important also to represent the places at which people can change mode. For example, a parking area represents a mode change from driving to walking or vice versa. Mode change locations supported in Commuter are as follows:

- Parking areas on-street, in bays, and off-street and in single-story or multistory lots;
- Public transit stands for buses, trams, and trains;
- Drop-off areas applying to private vehicles and taxis;
- Pick-up areas for private vehicles and prearranged collection points; and
- Taxi ranks.

Several interesting challenges were addressed concerning the layout of parking bays. Bays can be parallel to the direction of traffic or perpendicular, or in some cases they can be angled. It should also be possible for a vehicle to park on either side of the street and to leave a parking bay in either direction on a two-way street or aisle; this is valid especially in parking lots in which bays are perpendicular to traffic flow.

Objects Required to Model Person Trips

All people released into the model must have a place of origin. If the people in the model are to be directed, each person must also have a destination. These origin and destination areas are similar to the origin and destination zones commonly used in vehicle-oriented models, with the difference that the areas refer to a specific place that might not be accessible by vehicle. Defining a base and a ceiling for each area allows the definition of an area for one or more floors of a multistory building.

Once the origin and destination areas are defined, the model builds a routing decision "tree" to each possible destination, from every possible origin. This routing tree uses all available mode choice segments as its branches. There are branches created for all recognized modes of travel listed below. The routing tree is based on cost; each person has a behavior type that assigns costs to time, distance, and price.

Modes of Travel Recognized by the Model

The model must be able to represent several modes of travel for each person if the primary objective is to be achieved. These modes include

- Walking,
- Driving a private vehicle,
- Being a passenger in a private vehicle (having the state of "drop-off" or "pick-up"),
- Being a passenger in a fixed-route public transportation vehicle (bus, tram, train, etc.),
- Being a passenger in a taxi, and
- Waiting.

Of these, most are self-explanatory, apart from waiting, which can best be explained by example. Assume a person is traveling from home in a suburb to an office block in the city center. The office block does not have its own parking. The person has a car and can drive, and a park-and-ride facility is available at a nonwalkable distance from home. That person has the choice between

- Driving to a city center parking area and then walking to the office and
- Driving to the park-and-ride facility, walking from the parking bay to the train platform, waiting for the next train, riding the train to the city center, and then walking from the station to the office.

It is assumed that the person knows the timetable for the train, but in this example at least, the departure time from home is fixed, so it is inevitable that there will be some waiting time. It is important that explicit waiting time be included as a separate branch in the routing tree, or the total cost of traveling via the park-and-ride will not be correctly calculated. There are cost parameters for each mode, including waiting, so it is possible to assign a higher

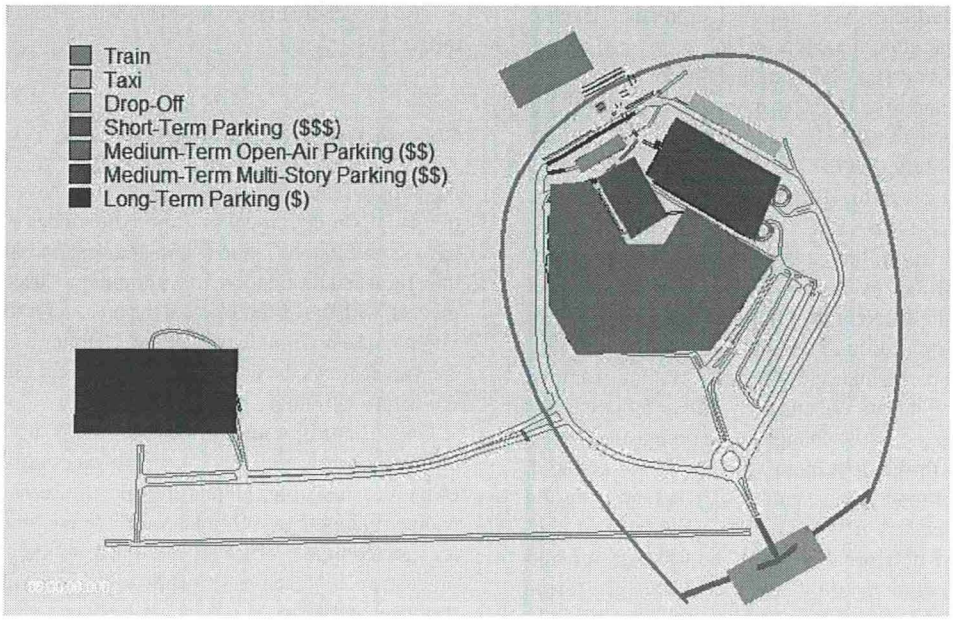


FIGURE 2 Overview of model layout showing access modes to airport terminal.

(or lower) cost to time spent waiting than to time spent riding on public transport.

Algorithms Used for Motion and Interaction of Agents

Many car-following models have been proposed for simulation, and each can be seen to have its advantages in certain situations (*11*). The software tool used for this work allows the user to select between three of the most popular models, referred to as Gipps (*12*), Wiedemann (*13*), and Fritszche (*14*) models. The Gipps model is used by default, being the simplest of the three to calibrate. Either of the other two can be selected by the user, with calibration parameters set accordingly, or it is possible for users to define their own car-following model by using the API.

The pedestrian motion uses a combined vector force model to steer each agent toward its target, avoiding obstacles and other agents (*15*). Although not as sophisticated as the models used in crowd modeling software packages, Commuter has not been designed for evacuation and other safety-related uses, so the simpler motion model appears to provide realistic patterns of movement. A range of parameters are provided for calibrating this model against observed pedestrian movements on road crossings and connecting walkways.

STUDY NETWORK

The network studied is based on Edinburgh Airport, United Kingdom. Some features of the network have been modified to compare as many access modes as possible. For example, an underground train has been added to the model, and the parking areas have been modified from those on the ground. The study network defines two O-D areas for people, one representing the city, the other representing the airport concourse for check-in and arrivals. Within this simple layout, there are seven possible modal combinations for the trip between city and airport, as shown in Figure 2. Six user types were defined, as shown in Table 2.

The cost values shown for each type are trial values only; calibration was not part of this research. Parking duration is used to calculate the dollar cost of each parking option, for each individual person. Each parking area has a cost calculated from a base cost plus an hourly rate multiplied by the number of hours for which a person will park his or her car. Each person entering the model calculates

TABLE 2 Person Types in Model, Showing Costs for Several Modes of Travel

Person Type	Parking Duration (h)	Eligible for Drop-Off	Can Park	Drive Time Cost (US$.01/s)	Drive Distance Cost (US$.01/km)	Walk Time Cost (US$.01/s)	Walk Distance Cost (US$.01/km)	Ride Time Cost (US$.01/s)	Ride Distance Cost (US$.01/km)	Time Cost (US$.01/s)
Business	16	No	Yes	0.03	0	0.5	1	2.5	50	2.5
Business	60	No	Yes	0.03	0	0.5	1	2.5	50	2.5
Business	120	No	Yes	0.03	0	0.5	1	2.5	50	2.5
Leisure	60	No	Yes	0.01	0	0.1	1	0.1	5	0.1
Leisure	180	No	No	0.01	0	0.1	1	0.1	5	0.1
Leisure	0	Yes	No	0.01	0	0.1	1	0.1	5	0.1

a total cost, in dollars, of each access method, based on the total cost of all mode segments of the trip required for that access method. For example, to calculate the cost of choosing to park in the long-term car park,

$$C = D_d w_{dd} + D_p w_{pd} + D_r w_{rd} + T_d w_{dt} + T_p w_{pt} + T_r w_{rt} + T_w w_{wt} + \frac{C_{\mathrm{LTB}} + T_p C_{\mathrm{LTH}}}{2}$$

where

$$D_d = D_{d0} + D_{d1}, T_d = T_{d0} + T_{d1}, D_p = D_{p1} + D_{p2}, T_p = T_{p1} + T_{p2}$$

D_{d0}, T_{d0} = distance and time of driving from the ultimate point of origin to the edge of the model;
D_{d1}, T_{d1} = distance and time of driving from the edge of the model to center of long-term parking;
D_{p1}, T_{p1} = distance and time of walking from the center of parking area to embarkation bus stop;
D_{p2}, T_{p2} = distance and time of walking from bus disembarkation stop to airport concourse;
D_r, T_r = distance and time of riding on the bus to the airport;
T_w = time to wait for the bus to the airport;
w_{dd}, w_{pd}, w_{rd} = cost-of-distance weights for driving, walking, and riding public transportation;
$w_{dt}, w_{pt}, w_{rt}, w_{wt}$ = cost-of-time weights for driving, walking, riding public transportation, and waiting;
$C_{\mathrm{LTB}}, C_{\mathrm{LTH}}$ = base and hourly price for parking; and
T_p = duration of stay in parking, in hours.

The total cost of parking is halved when the cost of the outward journey is calculated, to provide a fair comparison with the cost of other modes, in which single-fare information is used for cost. Alternatively, the entire cost of the parking could be used if that was compared with a return fare for taxi, train, and so forth.

The parameter settings for values of time used in the model attribute higher wealth to the business types, by placing a higher cost on time-consuming activities, such as walking. These values of time are commonly derived from stated preference surveys.

The screen views from the software shown in Figures 3, 4, and 5 illustrate some of the features from the model. Aerial photographs and 3-D models of buildings, cars, and people can be incorporated in the model, allowing even the nonexpert user to grasp the results of the analysis. People can travel in groups, such as family groups, or can push or pull an item of baggage, which requires more space in the model.

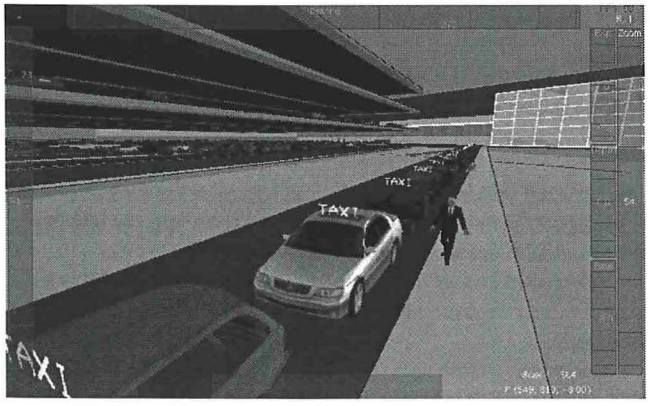


FIGURE 3 Taxi rank in the model where people change mode from walking to passenger.

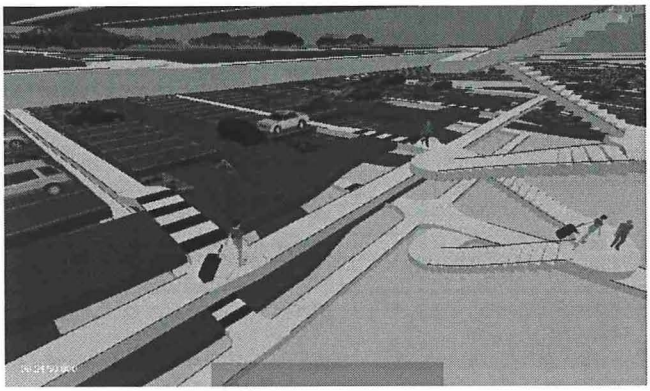


FIGURE 4 Parking area where people change mode from driving to walking and vice versa.

RESULTS

Performance: Speed and Cost

The model simulated a 4-h morning peak period (6:00 to 10:00 a.m.) in an average time of just under 25 min, achieving a speed-up of 9.6. The computer used to run the software was a dual-processor desktop PC purchased for approximately $1,500 in 2008.

Trips

A total of 5,335 trips were generated, consisting of 4,400 person trips (3,600 inward, 800 outward), 845 vehicle-only trips (used to simulate parking bays that were full at the beginning of the simulation period), and 90 public transit trips (41 train trips plus 49 bus trips). Of the trips that arrived before 10:00, the mean travel time from approach road to check-in was 4:37 (minutes:seconds), the median was 3:57, the minimum was 2:08, the lower quartile was 2:49, the upper quartile was 5:49, and the maximum was 8:42.

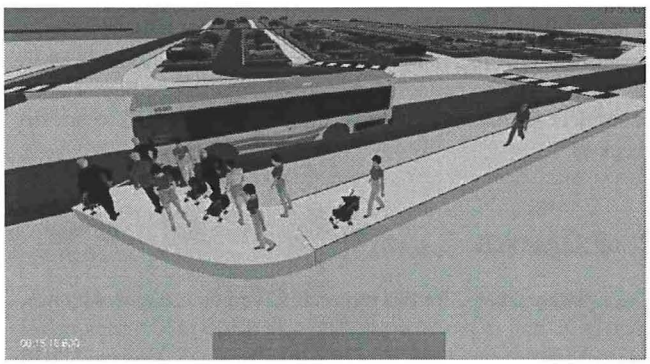


FIGURE 5 Shuttle bus stopping to pick up passengers from long-term parking area.

Efficient Methodology for Generating Synthetic Populations with Multiple Control Levels

Joshua Auld and Abolfazl Mohammadian

This paper details a new methodology for controlling attributes on multiple analysis levels in a population synthesis program. The methodology determines how household- and person-level characteristics can jointly be used as controls when populations are synthesized as well as how other multiple-level synthetic populations, such as firm and employee or household and vehicle, can be estimated. The use of multilevel controls is implemented through a new technique involving the estimation of household selection probabilities on the basis of the probability of observing each household, given the required person-level characteristics in each analysis zone. The new procedure is a quick and efficient method for generating synthetic populations that can accurately replicate desired person-level characteristics.

Population synthesis is recognized as an integral component in activity-based modeling. Beginning with the development of the TRANSIMS population synthesizer (*1*), increased focus has been directed at developing synthetic populations for use in travel demand microsimulation (*2–5*) and many other agent-based microsimulation applications (*6, 7*). Population synthesis generally uses a sample of households at an aggregate geography combined with marginal data on household characteristics at a disaggregate geography to generate a set of households that satisfy known marginals at the small-area level. Population synthesizers often use a well-known statistical technique, iterative proportional fitting (IPF) (*8*), and probabilistic selection to generate synthetic populations, although other procedures have recently been developed (*9*). Either way, a population synthesizer creates copies of sample households and locates them geographically to replicate the full population of the study area. For a more in-depth discussion of the IPF procedure and basic population synthesis techniques see Beckman et al. (*1*), Arentze et al. (*10*), and Hobeika (*11*), among others. The original population synthesis program in which the current work is implemented is discussed at length in Auld et al. (*12*). This program implemented the basic IPF procedure and probabilistic selection and was developed for use in an activity-based model system (*13*).

The increasing focus on population synthesis has resulted in recognition of some limitations of the basic synthesis method. This paper aims to improve the methodology behind the basic population synthesis routine to account for multiple levels of analysis units–control variables, which was a limitation to earlier population synthesizers. The paper includes a discussion of the literature on the issue, a description of a newly developed method to address it, validation of the new method and evaluations of its computational performance, and finally, a discussion of the value of the new method and directions for future work.

Department of Civil and Materials Engineering, University of Illinois at Chicago, 842 West Taylor Street, Chicago, IL 60607-7023. Corresponding author: J. Auld, auld@uic.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2175, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 138–147.
DOI: 10.3141/2175-16

PREVIOUS WORK IN POPULATION SYNTHESIS

The methodology behind most population synthesizers used in travel demand modeling is generally derived from the synthesizer developed by Beckman et al. (*1*) for the TRANSIMS project, although some recent work has also addressed the combinatorial optimization approach (*7, 9*) or combinations or permutations of both (*14, 15*). During the development of different population synthesizers, many limitations of the basic methodology have been observed. Subsequent research has focused on attempts to correct for these deficiencies and extend the usefulness of synthesis methods (*14, 16*). Several problematic issues relating to population synthesis that have been observed at various times include zero-cell issues arising from using sample data, biases introduced as a result of rounding the joint distributions, biases introduced as a result of simulation, and lack of multiple levels of control (*1, 9, 15, 16*). Different strategies have been proposed to address these issues, for example, the zero-cell problem has been addressed by tweaking the joint distribution from the IPF procedure (*1, 16*) and by limiting the number of control variable categories (*12, 16*).

The limitation of population synthesis methods to only one analysis level has recently begun to receive more attention. Traditionally, population synthesizers consider control variables for only one level because joint distributions between household- and person-level control variables cannot be constructed. Therefore the IPF procedure and selection procedure as found in Beckman et al. (*1*) cannot be implemented directly for household- and person-level variables simultaneously (*16*). Researchers have attempted to overcome this in several ways, including household reconstruction methods (*15*) or using population characteristics to impute household-level distributions (*10*). Recent work has focused on methods to address the issue directly in the synthesis procedure, rather than as a reconstruction step. Guo and Bhat account for person-level controls by developing joint distributions for individuals and households separately, and then synthesizing households while considering whether the person- or household-level constraints would be violated beyond a given

threshold, although only the household distribution is considered when drawing households (*16*). Ye et al. developed the only previous attempt of which the authors are aware to directly and simultaneously control on multiple levels (*14*). They used an iterative reweighting procedure to heuristically solve for household weights considering both household and person constraints together before the household selection procedure. The methodology presented here is a new, efficient procedure for considering joint multilevel controls implemented directly in the selection stage, which builds on the basic IPF and household draw procedure and which, to the best of the authors' knowledge, has not been implemented previously. For details of the basic procedure see Auld et al. (*12*). The new procedure is discussed in the following sections.

MULTILEVEL CONTROL METHODOLOGY

This section discusses the methodology used for multilevel control, implemented in the basic population synthesis program described in Auld et al. (*12*). Multilevel control allows population characteristics to be replicated when the synthetic population is created for more than one analysis level, with one level such as households serving as the base level of analysis that contains the sublevel analysis unit, such as persons. There is, however, no requirement that the analysis be used only for synthesizing households or individuals. Any situation in which marginal and sample data are available for a base- and sublevel of analysis (i.e., firms or employees, households or vehicles, buildings or tenants, etc.) can be synthesized by using the program. The only limitations are that the membership size of the sublevel within the base level must be used as a control (i.e., household size if using household or individual) and the sample data for the base- and sublevels must be linked by unique identifiers. The second requirement results from the fact that the program uses a procedure in which the base units are generated and their component subunits are copied with them rather than each subunit being synthesized separately. Because the subunits are copied with the base unit, there must be a link between the base- and subunit sample data. For clarity the base- and sublevels of analysis are referred to hereafter as simply household level and person level.

Household Selection Probability Considering Person-Level Constraints

One feature most population synthesizers share is the creation of synthetic households through probabilistic selection. This procedure involves setting a probability for selecting a sample household into the synthetic population on the basis of the sample weight of the household, the number of total households required, the number of households of the current type already generated, and so forth. This is the basic procedure followed in the synthesizer by Beckman et al. (*1*) and others. Selection probabilities are assigned for households that are then replicated through simulation. The probabilities increase with the weight of the household and decrease as the required frequency of the current household type is reduced through the simulation process. The required frequency of each household type is taken from the estimated household joint distribution created through the IPF process. Population synthesizers may depart from this basic methodology, as in the procedure developed by Ye et al., in which the frequencies determined in the IPF procedure are used in a heuristic iterative solution to set household weights such that person-level constraints are satisfied (*14*). Even in that case, however, simulation is still used to create the synthetic households by using the reweighted IPF results. The general selection probability as described in Beckman et al. is shown in Equation 1 (*1*).

$$P_{i,C} = \frac{W_i}{\sum_{k=1}^{N_C} W_k} \tag{1}$$

where

$P_{i,C}$ = probability of selecting household i of household type C,
W_i = household weight for household i, and
N_C = remaining households in subregion sample of type C.

This equation states that the probability of selecting the current household i of a given demographic type C is equal to the weight of the current household divided by the sum of the weights of all other households in the sample of the same type. This selection procedure ensures that households with a higher sample weight are selected more frequently when the households are synthesized. This selection probability does not account for differences between households on the person level. Therefore, a new selection probability, shown in Equation 2, was developed that explicitly accounts for the person-level distribution when the households are synthesized.

$$P_{i,C} = \frac{W_i \prod_{j=1}^{N_{\text{per},i}} \frac{\text{MWAY}^*_{\text{per}}\left(v_{1,j}, v_{2,j}, \ldots, v_{n,j}\right)}{N_{\text{remain}}}}{\sum_{k=1}^{N_C}\left(W_k \prod_{l=1}^{N_{\text{per},k}} \frac{\text{MWAY}^*_{\text{per}}\left(v_{1_l}, v_{2_l}, \ldots, v_{n_l}\right)}{N_{\text{remain}}}\right)} \tag{2}$$

where

$P_{i,C}$ = probability of selecting household i, of household type C,
W_i = household weight for household i,
$N_{\text{per},i}$ = number of people in household i,
$\text{MWAY}^*_{\text{per}}(v_{1,j}, \ldots, v_{n,j})$ = remaining cell frequency in zonal person-level joint distribution,
$v_{i,j}$ = index of control variable i for person j,
N_{remain} = number of individuals not yet created in zone, and
N_C = remaining households in subregion sample of type C.

The selection probability defined in Equation 2 has the same form as in Equation 1, with the addition of the product terms in the numerator and denominator. These product terms are essentially the probability of observing a household composed of each individual household member given the remaining persons to be synthesized according to the person-level joint distribution, $\text{MWAY}^*_{\text{per}}$. This selection probability is derived from a straightforward application of Bayes theorem, that is, the probability of selecting the current household H is the probability of observing household H given the current household type C. This is equivalent to the probability of observing each member in the household together divided by the sum of the probability of observing each household member together for all households of the same type, assuming no correlation between the probabilities for individual household members. This assumption is generally incorrect in actuality and would cause problems if households were

no households are selected no matter how many iterations are run. Therefore, on the final iteration of the procedure, if there are still households remaining to be generated, the program disregards all person-level controls and generates the remaining households on the basis of only the household weights by using the selection procedure seen in Equation 1.

PERSON-LEVEL CONTROL VALIDATION RESULTS

To assess the validity of the new person-level control methodology, a synthetic population created with the new routine was validated against the same population created without person-level control. The validation for the person-level control procedure was conducted on 846 block groups in the Chicago-land six-county region, where household- and person-level marginal control incompatibilities were minimal. Many block groups had populations less than the population estimated from the household size control variable, an error that causes less than the full number of households to be generated (because all person-level probabilities are set to zero before all households are generated). The selected block groups have a total of 553,387 households containing 1,498,482 individuals, approximately 20% of the total six-county population. These block groups were selected such that there were no group quarters population and the differences between estimated population totals based on the household size control variables and the population totals in the person-level marginals were less than 2%, to separate out error due to the procedure from error caused by data issues. Block groups with group quarters are excluded from this analysis only because including a marginal variable relating to group quarter status does not add anything to the person-level validation. When synthetic populations are generated for actual modeling purposes, it is a straightforward, although cumbersome, procedure to add a group quarters control marginal at the household level that enables block groups with substantial group quarters populations to be generated. In this manner, the validations run below are comparing the differences in procedure rather than differences due to data issues.

Two separate populations were synthesized, one by using only household controls referred to as POP-HH and one with an additional set of person-level controls referred to as POP-PER. The household controls used for both populations were

- Household size—seven categories,
- Household income—16 categories,
- Household number of workers—five categories, and
- Total household joint distribution size—560 cells.

The person-level controls used in generating POP-PER were

- Gender—two categories,
- Age—eight categories,
- Race—seven categories, and
- Total person joint distribution size—112 cells.

These variables were selected for demonstration purposes only; the purpose of this exercise is to confirm that using person-level controls improves the person-level fit results, not to validate the use of this particular set of control variables. Any set of household- and person-level variables for which adequate sample and marginal data exist can be used because the synthesis program is designed to be as general as possible (*12*).

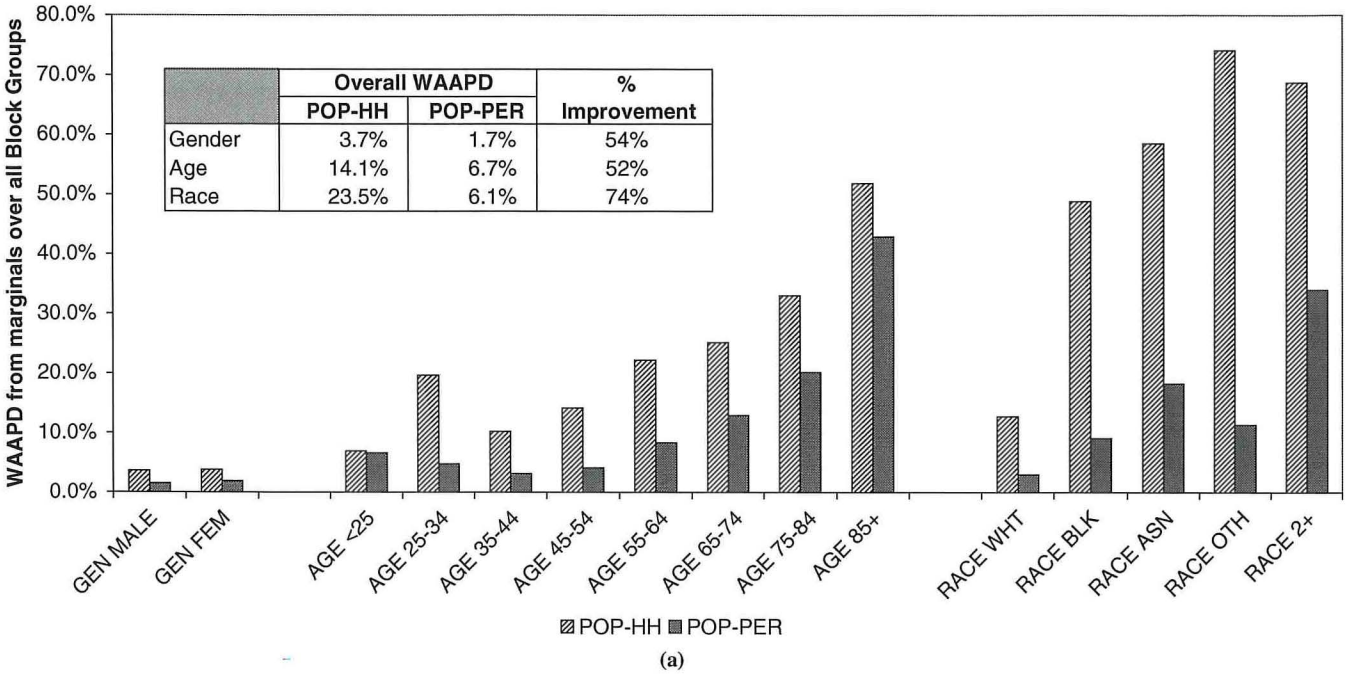
Both synthetic populations were able to exactly match the total number of households required, with each generating the actual total of 553,387 households. In addition the total number of individuals generated was almost exact for each synthetic population, as expected even for the non-person control population as a result of the inclusion of a household size variable as a control. The POP-HH population contains 1,500,308 people, 0.1% more than required, whereas the POP-PER population contains 1,487,815 people, 0.7% less than required. The marginal fit comparison, in regard to weighted average absolute percent difference (WAAPD) between the known and synthesized marginal totals over all block groups, for both populations is shown in Figure 2. The Native American, Alaskan, and Hawaiian categories in the race control are not shown because these categories represent less than 0.25% of the population in the region although both exhibited improvement similar to the other categories.

The person-level comparison, shown in Figure 2*a,* demonstrates a substantial improvement in fit between the POP-HH and POP-PER marginal totals on the person level, as expected. Overall there is an improvement in fit of between 52% and 74% over each person-level category, showing that the new routine allows a marked improvement in fitting to person-level marginal control totals. As seen in the figure, even under person-level control, the average error associated with certain marginal categories can still be large, although always less than with no person control. This is the result mainly of rounding errors and difficulty satisfying the marginal constraints for infrequent categories. The largest errors in the marginal fit are seen for the over-85-years-of-age category and the two-or-more-races category for the age and race marginals, respectively, which each represent less than 2% of the total population. In fact all marginal categories that have a WAAPD of more than 15% contain less than 5% of the population, meaning that the large errors are the result mainly of small category sizes.

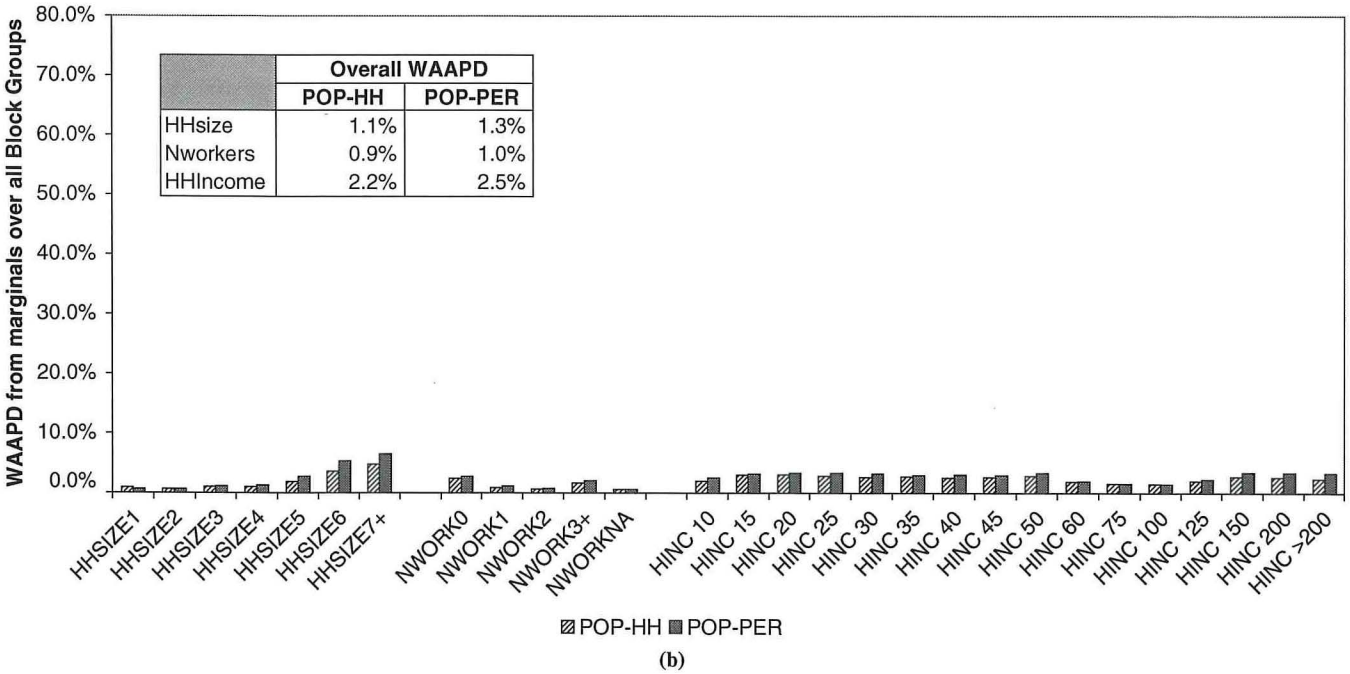
The household-level comparison in Figure 2*b* shows that the improvement in marginal fit using person-level controls comes at a minimal cost to the accuracy of the household-level marginals. All marginal control totals are matched fairly precisely in the POP-HH and POP-PER synthetic populations, with larger errors again seen in the less frequent categories. All household marginal categories had under a 7.0% WAAPD value.

One point about the procedure should be noted concerning the relaxation of the person-level constraints used to ensure convergence when selecting households. It is clear that allowing the person-level constraints to be violated introduces errors into matching the expected person-level marginals, causing most of the differences seen in Figure 2*a.* However, analysis shows that in general it is a very small number of generated households and individuals that contribute to these violations, so the impacts are most likely not particularly large. For the POP-PER synthetic population, on average more than 97% of households (2% standard deviation) and 95% of individuals (3% standard deviation) were generated before the person-level constraints were relaxed.

The previous analysis shows only how the population matches the marginal characteristics. Therefore each synthetic population was also evaluated on how well the required household- and person-level joint distributions were matched. This is evaluated by estimating the absolute percent difference between the synthesized and expected (from IPF) frequencies for each cell in each block group. This value is then averaged over all block groups to obtain an average absolute percent difference (AAPD) value for each cell in each joint distribution. The AAPD values for each synthetic population are then plotted against the average cell frequency, along with a



(a)



(b)

FIGURE 2 WAAPD comparison for (*a*) person-level marginals and (*b*) household-level marginals.

theoretical estimated AAPD from rounding error calculated as shown in Equation 3 below:

$$\mathrm{AAPD}_i = \frac{\sum_{j=1}^{N_{\mathrm{BG}}} \mathrm{APD}_{i,j}}{N_{\mathrm{BG}}}$$

$$\mathrm{APD}_{i,j} = (1-p_{i,j})\frac{x_{i,j}-(x_{i,j}-p_{i,j})}{x_{i,j}} + p_{i,j}\frac{(x_{i,j}+1-p_{i,j})-x_{i,j}}{x_{i,j}}$$

$$= \frac{2(p_{i,j})(1-p_{i,j})}{x_{i,j}} \quad (3)$$

where

$\mathrm{APD}_{i,j}$ = expected absolute percentage difference from value in cell i for block group j from rounding,
AAPD_i = average APD for cell i over all block groups from rounding,
$p_{i,j} = x_{i,j} \pmod 1$,
$x_{i,j}$ = value in cell i of person-level joint distribution for blockgroup j, and
N_{BG} = number of block groups (zones).

Equation 3 states that the expected absolute percent difference for each cell in the joint distribution for each block group is the probability of rounding the cell down multiplied by the error caused by this plus the probability of rounding the cell up multiplied by the error caused from rounding up, where the probability is determined by the decimal portion of the actual cell value (i.e., a cell value of 1.2 will be rounded down 80% of the time and rounded up 20% of the time, so that 80% of the time the error is 0.2/1.2, or 16.7%, and 20% of the time the error is 0.8/12, or 67%, for an average of 26.7%). The values for each block group are then averaged to obtain the AAPD value for each cell. These values are plotted, along with the AAPD values from the POP-HH and POP-PER populations in Figure 2 for the household- and person-level joint distributions. Note that these values are plotted against average cell frequency, so that a cell with an integer average frequency will still have expected average rounding error.

Figure 3*a* shows the results of the comparisons of the AAPD values for each cell in the household distribution matrix for the POP-HH and POP-PER synthetic populations. The figure shows that the populations produced through both procedures replicate the household-level joint distribution reasonably well, with the AAPD values approaching the theoretically expected value as a result of random rounding. In fact, the population generated with person controls actually slightly outperforms the base procedure in satisfying the household distribution with an average AAPD over all cells of 89% compared with 125% for the POP-HH population. That is possibly the result of a more targeted search being performed through the use of the person-level controls and constraints.

The results presented in Figure 3*b* show that, as expected, the fit of the POP-PER synthetic population to the person-level joint distribution is much better than the fit of the POP-HH population, as a result of the use of the person-level controls. The overall AAPD improves from 407% for the POP-HH to 118% for the POP-PER population, which is a significant improvement. The cell AAPD values for the POP-PER population are generally much closer to the expected rounding error, whereas large differences can be seen in the POP-HH AAPD. Although the POP-PER AAPD values also generally follow the expected pattern of decreasing error with increasing average cell size, that is not the case with the uncontrolled population, with large errors seen even for several cells with large average sizes, which reinforces the problem with not controlling for person-level characteristics. This result is not due merely to the error caused by large variances in the household size between zones because that is accounted for in the calculation of the expected AAPD value.

Overall, the validation analyses presented in Figures 2 and 3 show that the additional use of person-level controls when a population is generated improves the fit of the resulting population to known person-level characteristics when compared with the same synthetic population generated without person-level controls. The increase in fit to the person-level known marginal totals and estimated joint distribution is very substantial, with little to no sacrifice in the ability to match household-level characteristics. In fact, the ability to match the household joint distribution is somewhat improved through the use of the person-level controls.

A final validation exercise was performed to determine whether the new, more-efficient selection procedure outlined in the section on the updated household selection procedure had any negative impact on the fit of the synthetic populations when compared with the traditional selection procedure. For this validation analysis the selection procedure refers only to the manner in which the sample households are searched; both procedures tested here still use the new household selection probability calculation, which accounts for person-level characteristics. Also, because the test is conducted to determine the validity of the selection procedure rather than the overall synthesis procedure, the marginal constraints were turned off when the test synthetic populations were generated. Three different synthetic populations were generated for 46 block groups in Public Use Microdata Areas (PUMAs) 3408, 3409, 3518, and 3519 in the Chicago region, which had no group quarters population and minimal discrepancies between household size counts and population levels. The three populations were person-level control under the new selection procedure (PER-NEW), person-level control under the traditional selection procedure (PER-OLD), and no person control (PER-NONE).

To test for potential biases in the new selection procedure, the Freeman–Tukey test statistic was used to compare the fit of the generated household and person joint distributions with the expected distributions from the IPF process for each procedure. The advantages of this statistic for use in analyzing goodness of fit for synthetic population have been described in Voas and Williamson (*17*) and Ryan et al. (*7*). The test statistic is calculated as

$$\mathrm{FT}^2 = 4\sum_{i}^{N_{\mathrm{cells}}}\sum_{j}^{N_{\mathrm{zones}}}\left(\sqrt{\hat{u}_{ij}}-\sqrt{u_{ij}}\right)^2$$

$$\mathrm{FT}^2 \sim \chi^2\left(N_{\mathrm{cells}} \times N_{\mathrm{zones}} - 1\right) \quad (4)$$

where the statistic is four times the sum of the square of the differences between the square root of actual (u_{ij}) and estimated ($\hat{u}_{ij}$) frequencies over all cells i and zones j and has a chi-square distribution. The test statistic is calculated and compared with a critical value for a given significance level from the χ^2 distribution to evaluate the fit of the synthesized population to the person-level joint distribution. The results for all three synthetic populations are shown in Table 1 for the household- and person-level distributions at a significance level of .05.

According to Table 1, the null hypothesis for the Freeman–Tukey test—that is, the synthesized joint distribution and joint distribution resulting from IPF at the person level have the same distribution—

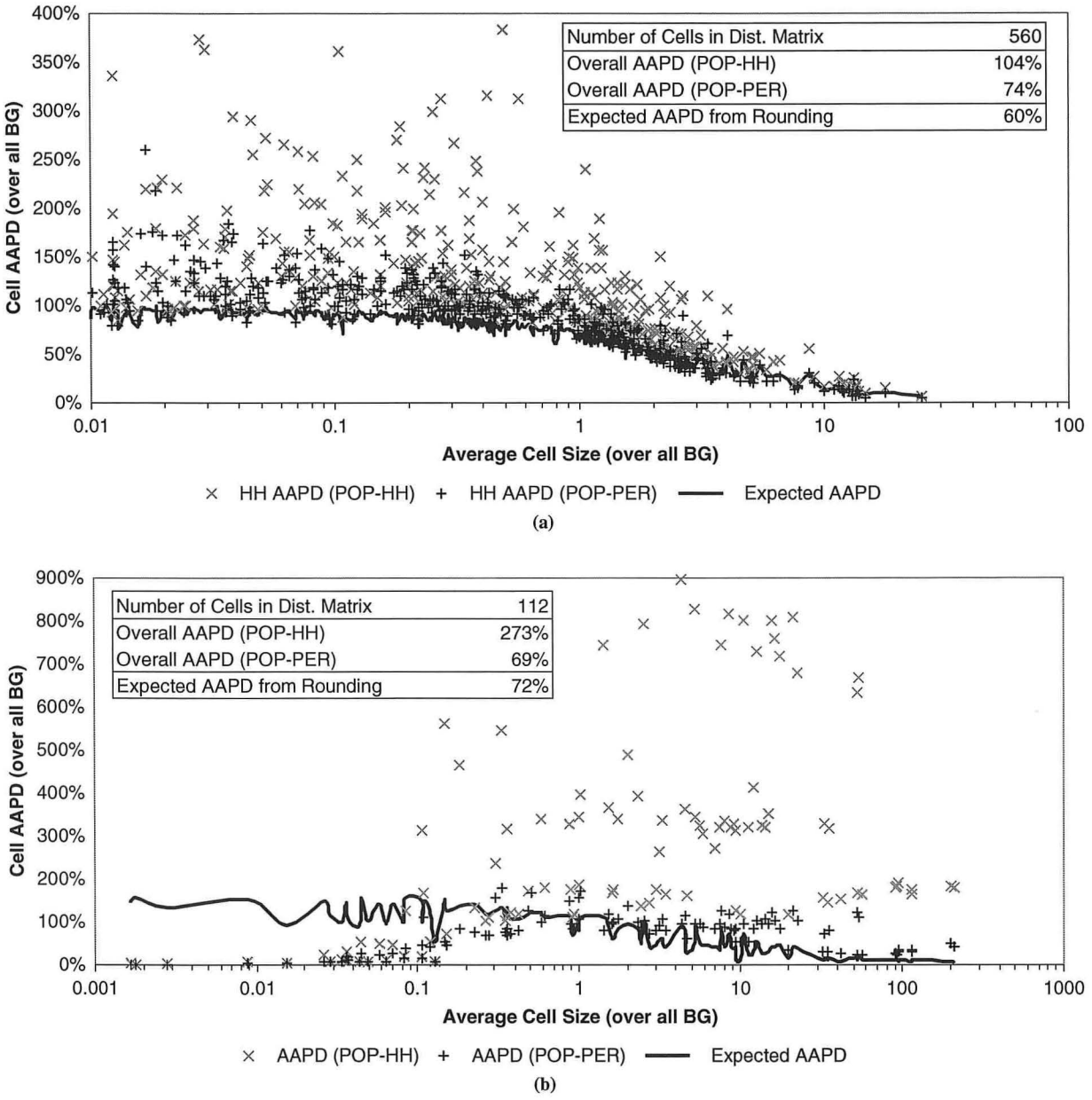


FIGURE 3 AAPD comparison for (*a*) household-level joint distribution and (*b*) person-level joint distribution.

TABLE 1 Comparison of Synthetic Population Fit for Different Selection Procedures

	Household-Level Distribution[a,c]			Person-Level Distribution[b,c]		
Population	Crit Value	FT^2 (σ)	H_o^d	Crit Value	FT^2 (σ)	H_o^d
PER-NONE	26,134	4,799 (54)	Accept	5,319	24,786 (434)	Reject
PER-OLD	26,134	5,734 (68)	Accept	5,319	4,044 (106)	Accept
PER-NEW	26,134	6,651 (82)	Accept	5,319	4,840 (102)	Accept

[a]25,759 degrees of freedom for household-level distribution.
[b]5,151 degrees of freedom for person-level distribution.
[c]FT^2 values averaged over 20 runs; standard deviation of FT^2 value shown in parentheses.
[d]Null hypothesis accepted if FT^2 is less than critical value at significance level of .05, i.e., probability of observing FT^2 statistic due to random chance is greater than 5%.

is accepted for both populations with person-level controls and rejected for the population without controls, and the household-level distribution is matched for all populations. The results in Table 1 clearly show that using person-level controls improves the fit of the synthesized person-level joint distribution to the estimated distribution, whereas not controlling for person-level characteristics results in poor fit to the estimated distribution, as expected. More important, the good fit to the joint distribution is obtained for both selection procedures. Although the fit obtained by using the new procedure is slightly worse than that obtained by using the traditional procedure, it is still good and results in a run time of 0.7 min to synthesize the 85,590 individuals in the example above as compared with 18.6 min using the other procedure. The run time for synthesizing the entire population in the Chicago region with the traditional procedure assuming the same rates obtained above would be approximately 30 h for a single run compared with the 1.4 h achieved with the new procedure. The long run times using the traditional selection procedure combined with the potential need for running multiple different permutations of a synthetic population and for averaging over multiple runs for the same population motivates the use of the more efficient selection procedure, although the traditional selection procedure can still be used to generate a final synthetic population in combination with initial testing and development done by using the faster procedure. For this reason, both selection procedures are implemented in the actual synthesis program with the choice left to the user.

COMPUTATION PERFORMANCE

Beyond validating the accuracy of the new methodology, it is necessary to evaluate its computation performance. To determine the performance characteristics of the new algorithm, the run times for generating the synthetic populations described in the previous section, POP-HH and POP-PER were compared with run times for generating the full Chicago population, with and without person-level controls. The same program settings, other than the use of person control, were used in each run. Each synthetic population was generated by running the population synthesis program on an Intel Centrino Duo 2.0-GHz processor.

The non-person-controlled population, POP-HH, which contained 1,500,308 synthetic individuals, took 13 min to generate. In contrast, the population with person-level controls, POP-PER, with 1,487,815 people, took more than 28 min. For the full populations, the non-person-controlled full population took about 33 min to generate 7,972,057 individuals, and the person-controlled full population took 84 min to generate 7,889,221, out of a total actual population of 8,091,720. All of the synthetic populations had a household-level joint distribution size of 560 cells and a person-level joint distribution size of 112 cells.

Although it is difficult to compare results across different synthesizers, these run times appear to compare favorably as far as the authors can tell. During the validation of the Atlanta Regional Council population synthesizer, a synthetic population of 1.35 million households controlled only at the household level was run in 17.4 min with a household-distribution size of 316 cells, about half the time it took to synthesize the 2.9 million households in the Chicago region by using only household controls in the new synthesizer (*18*).

The only comparable results available for synthesizers that control for person-level characteristics were presented in Ye et al. for a synthetic population of 2.9 million individuals in Maricopa County, Arizona (*14*). This synthetic population was generated by using a household-distribution size of 280 cells (more than three control variables) and a person-joint-distribution size of 140 cells (over the same three control variables used in this study but with two additional age categories). The overall run time was 16 h, which is substantially longer than the 1.4 h to generate the Chicago population of 7.9 million individuals with approximately the same number of control variables and distribution matrix sizes.

CONCLUSIONS

This paper has detailed the development of a new methodology for using control variables at multiple analysis levels when synthesizing populations with an existing population synthesizer (*12*). The new procedure improves the fit of the synthesized person-level characteristics when compared with synthesis procedures that do not account for person-level controls. Validation of the new methodology shows that the improved fit to the person controls comes at no cost to the fit against the household-level controls. In addition, the introduction of a new household selection procedure has greatly increased efficiency while maintaining good fit to the required person-level controls without some of the run time issues that are found with the use of other methods. Although the discussion in this paper is limited mainly to the household/person synthesis, this methodology can be applied to any analysis with multiple levels of control. Future work is expected on generating shipping firms/vehicles and business firms/employees, for example, by using the same synthesis program. In fact, the applicability of the program is limited only by the availability of data. Overall, the new methodology seems to be an improvement on existing population synthesis techniques for controlling characteristics on multiple levels of analysis.

ACKNOWLEDGMENTS

The authors acknowledge the financial support provided to this project from the Chicago Metropolitan Agency for Planning and the National Science Foundation Integrative Graduate Education and Research Traineeship Program in Computational Transportation Science at the University of Illinois at Chicago.

REFERENCES

1. Beckman, R. J., K. A. Baggerly, and M. D. McKay. Creating Synthetic Baseline Populations. *Transportation Research Part A,* Vol. 30, No. 6, 1996, pp. 415–429.
2. Roorda, M. J., E. J. Miller, and K. Habib. Validation of TASHA: A 24-Hour Activity Scheduling Microsimulation Model. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C., 2007.
3. Bhat, C. R., J. Y. Guo, S. Srinivasan, and A. Sivakumar. Comprehensive Econometric Microsimulator for Daily Activity-Travel Patterns. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1894,* Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 57–74.
4. Yagi, S., and A. Mohammadian. Modeling Daily Activity–Travel Tour Patterns Incorporating Activity Scheduling Decision Rules. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2076,* Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 123–131.
5. Frick, M., and K. W. Axhausen. Generating Synthetic Populations Using IPF and Monte Carlo Techniques: Some New Results. Presented at 4th Swiss Transport Research Conference, Monte Verità, Switzerland, March 25–26, 2004.

6. Wheaton, W. D., J. C. Cajka, B. M. Chasteen, D. K. Wagener, P. C. Cooley, L. Ganapathi, et al. *Synthesized Population Databases: A U.S. Geospatial Database for Agent-Based Models.* RTI Press Publication No. MR-0010-0905. RTI Press, Research Triangle Park, N.C., 2009.
7. Ryan, J., H. Moah, and P. Kanaroglou. Population Synthesis: Comparing the Major Techniques Using a Small, Complete Population of Firms. *Geographical Analysis,* Vol. 41, No. 2, 2009, pp. 181–203.
8. Deming, W. E., and F. F. Stephan. On a Least Squares Adjustment of a Sampled Frequency When the Expected Marginal Totals Are Known. *Annals of Mathematical Statistics,* Vol. 11, 1940, pp. 427–444.
9. Voas, D., and P. Williamson. An Evaluation of the Combinatorial Optimization Approach to the Creation of Synthetic Microdata. *International Journal of Population Geography,* Vol. 6, No. 5, 2000, pp. 349–366.
10. Arentze, T., H. J. P. Timmermans, and F. Hofman. Creating Synthetic Household Populations: Problems and Approach. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2014,* Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 85–91.
11. Hobeika, A. *TRANSIMS Fundamentals: Chapter 3 Population Synthesizer, Technical Report.* U.S. Department of Transportation. Washington, D.C., July 2005. http://tmip.fhwa.dot.gov/transims/transims_fundamentals/ch3.pdf. Accessed Aug. 1, 2007.
12. Auld, J. A., A. Mohammadian, and K. Wies. Population Synthesis with Subregion-Level Control Variable Aggregation. *ASCE Journal of Transportation Engineering,* Vol. 135, No. 9, 2009.
13. Auld, J. A., and A. Mohammadian. Framework for the Development of the Agent-Based Dynamic Activity Planning and Travel Scheduling Model. *Transportation Letters: The International Journal of Transportation Research,* Vol. 1, No. 3, 2009, pp. 243–253.
14. Ye, X., K. C. Konduri, R. M. Pendyala, B. Sana, and P. Waddell. Methodology to Match Distributions of Both Household and Person Attributes in Generation of Synthetic Populations. Presented at 88th Annual Meeting of the Transportation Research Board, Washington, D.C., 2009.
15. Pritchard, D. R., and E. J. Miller. Advances in Agent Population Synthesis and Application in an Integrated Land Use and Transportation Model. Presented at 88th Annual Meeting of the Transportation Research Board, Washington, D.C., 2009.
16. Guo, J. Y., and C. R. Bhat. Population Synthesis for Microsimulating Travel Behavior. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2014,* Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 92–101.
17. Voas, D., and P. Williamson. Evaluating Goodness-of-Fit Measures for Synthetic Microdata. *Geographical and Environmental Modeling,* Vol. 5, No. 2, 2001, pp. 177–200.
18. Bowman, J. L., and G. Rousseau. Validation of the Atlanta Population Synthesizer. Prepared for TRB Conference on Innovations in Travel Modeling, Austin, Tex., May 21–23, 2006.

The Transportation Demand Forecasting Committee peer-reviewed this paper.

TRANSPORTATION RESEARCH RECORD:
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Peer Review Process

The *Transportation Research Record: Journal of the Transportation Research Board* publishes approximately 25% of the more than 3,000 papers that are peer reviewed each year. The mission of the Transportation Research Board (TRB) is to disseminate research results to the transportation community. The Record series contains applied and theoretical research results as well as papers on research implementation.

The TRB peer review process for the publication of papers allows a minimum of 30 days for initial review and 60 days for rereview, to ensure that only the highest-quality papers are published. At least three reviews must support a committee's recommendation for publication. The process also allows for scholarly discussion of any paper scheduled for publication, along with an author-prepared closure.

The basic elements of the rigorous peer review of papers submitted to TRB for publication are described below.

Paper Submittal: June 1–August 1

Papers may be submitted to TRB at any time. However, most authors use the TRB web-based electronic submission process available between June 1 and August 1, for publication in the following year's Record series.

Initial Review: August 15–November 15

TRB staff assigns each paper by technical content to a committee that administers the peer review. The committee chair assigns at least three knowledgeable reviewers to each paper. The initial review is completed by mid-September.

By October 1, committee chairs make a preliminary recommendation, placing each paper in one of the following categories:

1. Publish as submitted or with minor revisions,
2. Publish pending author changes and rereview, or
3. Reject for publication.

By late October, TRB communicates the results of the initial review to the corresponding author indicated on the paper submission form. Corresponding authors communicate the information to coauthors. Authors of papers in Category 2 (above) must submit a revised version addressing all reviewer comments and must include a cover letter explaining how the concerns have been addressed.

Rereview: November 20–January 25

The committee chair reviews revised papers in Category 1 (above) to ensure that the changes are made and sends the Category 2 revised papers to the initial reviewers for rereview. After rereview, the chairs make the final recommendation on papers in Categories 1 and 2. If the paper has been revised to the committee's satisfaction, the chair will recommend publication. The chair communicates the results of the rereview to the authors.

Discussions and Closures: February 1–May 15

Discussions may be submitted for papers that will be published. TRB policy is to publish the paper, the discussion, and the author's closure in the same Record.

Many papers considered for publication in the *Transportation Research Record* are also considered for presentation at TRB meetings. Individuals interested in submitting a discussion of any paper presented at a TRB meeting must notify TRB no later than February 1. If the paper has been recommended for publication in the *Transportation Research Record*, the discussion must be submitted to TRB no later than April 15. A copy of this communication is sent to the author and the committee chair.

The committee chair reviews the discussion for appropriateness and asks the author to prepare a closure to be submitted to TRB by May 15. The committee chair reviews the closure for appropriateness. After the committee chair approves both discussion and closure, the paper, the discussion, and the closure are included for publication together in the same Record.

Final Manuscript Submittal: March 15

In early February, TRB requests a final manuscript for publication—to be submitted by March 15—or informs the author that the paper has not been accepted for publication. All accepted papers are published by December 31.

Paper Awards: April to January

The TRB Executive Committee has authorized annual awards sponsored by Groups in the Technical Activities Division for outstanding published papers:

- Charley V. Wootan Award (Policy and Organization Group);
- Pyke Johnson Award (Planning and Environment Group);
- K. B. Woods Award (Design and Construction Group);
- Patricia F. Waller Award (Safety and System Users Group);
- D. Grant Mickle Award (Operations and Preservation Group); and
- John C. Vance Award (Legal Resources Group).

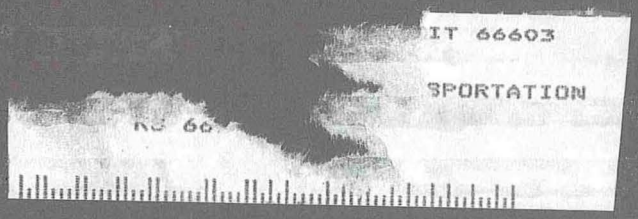
Other Groups also may nominate published papers for any of the awards above. In addition, each Group may present a Fred Burggraf Award to authors 35 years of age or younger.

Peer reviewers are asked to identify papers worthy of award consideration. Each Group reviews all papers nominated for awards and makes a recommendation to TRB by September 1. TRB notifies winners of the awards, which are presented at the following TRB Annual Meeting.

Transportation Research Board
www.TRB.org

TRANSPORTATION RESEARCH BOARD
500 Fifth Street, NW
Washington, DC 20001

ADDRESS SERVICE REQUESTED



THE NATIONAL ACADEMIES™
Advisers to the Nation on Science, Engineering, and Medicine

The nation turns to the National Academies—National Academy of Sciences, National Academy of Engineering, Institute of Medicine, and National Research Council—for independent, objective advice on issues that affect people's lives worldwide.
www.national-academies.org

ISBN 978-0-309-16051-3
90000
9 780309 160513